Linear-Rational Term Structure Models *

Damir Filipović  Martin Larsson  Anders Trolle
EPFL and Swiss Finance Institute
September 4, 2014

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 nonnegative interest rates, ii) easily accommodates unspanned factors affecting volatility and risk premiums, and iii) admits analytical solutions to swaptions. A parsimonious model specification within the linear-rational class has a very good fit to both interest rate swaps and swaptions since 1997 and captures many features of term structure, volatility, and risk premium dynamics—including when interest rates are close to the zero lower bound.

Keywords: Swaps, Swaptions, Unspanned Factors, Zero Lower Bound

JEL Classification: E43, G12, G13

*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, the Conference of Current Topics in Mathematical Finance in Vienna, the Princeton-Lausanne Workshop on Quantitative Finance in Princeton, the 29th European Meeting of Statisticians in Budapest, the Term Structure Modeling at the Zero Lower Bound workshop at the Federal Reserve Bank of San Francisco, the Cambridge Finance Seminar, the London Mathematical Finance Seminar, the Seminar on Mathematical Finance in Vienna, the Stanford Financial and Risk Analytics Seminar, the Symposium on Interest Rate Models in a Low Rate Environment at Claremont Graduate University, the 2014 Quant-Europe conference in London, the 2014 Global Derivatives conference in Amsterdam, the Federal Reserve Bank of New York, Copenhagen Business School, and Andrew Cairns, Pierre Collin-Dufresne, Darrell Duffie, and Ken Singleton (discussant) 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.
1 Introduction

The current environment with very low interest rates creates difficulties for many existing term structure models, most notably Gaussian or conditionally Gaussian models that invariably place large probabilities on negative future interest rates. Models that respect the zero lower bound (ZLB) on interest rates exist but are often restricted in terms of accommodating unspanned factors affecting volatility and risk premiums and in terms of pricing many types of 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 nonnegative interest rates, ii) easily accommodates unspanned factors affecting volatility and risk premiums, 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 of a set of parsimonious model specifications within the linear-rational class.

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 $\zeta_t$ such that the price $\Pi(t, T)$ at time $t$ of any time $T$ cash-flow, $C_T$ say, is given by

$$\Pi(t, T) = \frac{1}{\zeta_t} E_t[\zeta_T C_T].$$

(1)

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 linear drift, and a state price density, which is a linear function of the current state. In this case, bond prices and the short rate become linear-rational functions—i.e., ratios of linear 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 nonnegative interest rates. Another attractive feature is that the martingale component of the factor process does not affect the term structure. Therefore, one can easily allow for factors that affect prices of interest rate derivatives without affecting bond prices. Assuming that the factor

---

1 Throughout, we assume there is a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ on which all random quantities are defined, and $\mathbb{E}_t[\cdot]$ denotes $\mathcal{F}_t$-conditional expectation.

2 While zero appears to be a sensible lower bound on nominal interest rates, any lower bound is accommodated by the framework.
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 the distribution of future bond prices. 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 both swaps and swaptions from January 1997 to August 2013. At a weekly frequency, we observe a term structure of swap rates with maturities from 1 year to 10 years as well as a surface of at-the-money implied volatilities of swaptions with swap maturities from 1 year to 10 years and option expiries from 3 months to 5 years. The estimation approach is quasi-maximum likelihood in conjunction with the Kalman filter. The term structure is assumed to be driven by three factors, and we vary the number of USV factors between one and three. The preferred specification has three USV factors and has a very good fit to both swaps and swaptions simultaneously. This holds true also for the part of the sample period when short-term rates are very close to the ZLB.

Using long samples of simulated data, we investigate the ability of the model to capture the dynamics of the term structure, volatility, and swap risk premiums. First, the model captures important features of term structure dynamics near the ZLB. Consistent with the data, the model generates extended periods of very low short rates. Furthermore, when the short rate is close to zero, the model generates highly asymmetric distributions of future short rates, with the most likely values of future short rates being significantly lower than the mean values. Related to this, the model replicates how the first principal component of the term structure changes from being a “level” factor during normal times to being more of a “slope” factor during times of near-zero short rates.

Second, the model captures important features of volatility dynamics near the ZLB. Previous research has shown that a large fraction of the variation in volatility is largely unrelated to variation in the term structure. We provide an important qualification to this result: volatility becomes compressed and gradually more level-dependent as interest rates approach the ZLB. This is illustrated by Figure 1, which shows the 3-month implied volatility of the 1-year swap rate plotted against the level of the 1-year swap rate. More formally, for each swap maturity, we regress weekly changes in the 3-month implied volatility of the swap rate on weekly changes in the level of the swap rate. Conditional on swap rates being close to
zero, the regression coefficients are positive, large in magnitudes, and very highly statistically significant, and the $R^2$s are around 0.50. However, as the level of swap rates increases, the relation between volatility and swap rate changes becomes progressively weaker, and volatility exhibits very little level-dependence at moderate levels of swap rates. Capturing these dynamics—strong level dependence of volatility near the ZLB and predominantly USV at higher interest rate levels—poses a significant challenge for existing dynamic term structure models. Our model successfully meets this challenge because it simultaneously respects the ZLB on interest rates and incorporates USV.

Third, the model captures several characteristics of swap risk premiums. We consider realized excess returns on swap contracts (from the perspective of the party who receives fixed and pays floating) and show that in the data the unconditional mean and volatility of excess returns increase with swap maturity, but in such a way that the unconditional Sharpe ratio decreases with swap maturity.\(^3\) We also find that implied volatility is a robust predictor of excess returns, while the predictive power of the slope of the term structure is relatively weak in our sample.\(^4\) The model largely captures unconditional risk premiums and, as the dimension of USV increases, has a reasonable fit to conditional risk premiums.\(^5\)

The linear-rational framework is related to the linearity-generating (LG) framework studied in Gabaix (2009) and Cheridito and Gabaix (2008) in which bond prices are linear functions of the current state.\(^6\) Indeed, we show that term structure models based on LG processes are strictly included in the linear-rational framework. A specific LG term structure model is analyzed by Carr, Gabaix, and Wu (2009); however, the factor process in their model is time-inhomogeneous and non-stationary, while the one in the LRSQ model is time-homogeneous and stationary. Also, the volatility structure is very different in the two models, and bond prices are perfectly correlated in the Carr, Gabaix, and Wu (2009) model, while bond prices exhibit a truly multi-factor structure in the LRSQ model.

The exponential-affine framework, see, e.g., Duffie and Kan (1996) and Dai and Singleton

\(^3\)This is similar to the findings of Duffee (2010) and Frazzini and Pedersen (2014) in the Treasury market.

\(^4\)This result differs from a large literature on predictability of excess bond returns in the Treasury market. The reason is likely some combination of our more recent sample period, our use of forward-looking implied volatilities, and structural differences between the Treasury and swap markets. As we note later, a key property of many equilibrium term structure models is a positive risk-return trade-off in the bond market, which is consistent with our results.

\(^5\)The historical mean excess returns and Sharpe ratios are inflated by the downward trend in interest rates over the sample period and, indeed, the model-implied values are lower.

\(^6\)More generally, the linear-rational framework is related to the frameworks in Rogers (1997) and Flesaker and Hughston (1996).
(2000), is arguably the dominant one in the term structure literature. In this framework, one can either ensure nonnegative interest rates (which requires all factors to be of the square-root type) or accommodate USV (which requires at least one conditionally Gaussian factor, see Joslin (2014)), but not both. Furthermore, no exponential-affine model admits analytical solutions to swaptions.\footnote{Various approximate schemes for pricing swaptions have been proposed in the literature; see, e.g., Singleton and Umantsev (2002) and the references therein.} In contrast, the linear-rational framework accommodates all three features.\footnote{Alternative frameworks that ensure nonnegative interest rates include the “shadow rate” framework of Black (1995) (see, e.g., Kim and Singleton (2012) for a recent application) and the exponential-quadratic framework of Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002). Neither of these frameworks accommodate USV or admit analytical solutions to swaptions.} We compare the LRSQ model with the exponential-affine model that is based on a multi-factor square-root process. Because of the limitations of the latter model, we abstract from USV and swaption pricing and focus exclusively on the pricing of swaps. The two models have the same total number of parameters, but the LRSQ model has fewer parameters affecting the term structure since swap rates do not depend on the diffusion parameters. Nevertheless, the two models perform virtually identically in terms of pricing swaps, while the LRSQ model has a better fit to factor volatilities because the identification of the diffusion parameters only comes from the time-series of the factors. Comparing the short-rate dynamics in the two models, we find that the LRSQ model exhibits a moderate degree of nonlinearity in both the drift and instantaneous variance of the short rate.

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 discusses a flexible specification of market prices of risk. The empirical analysis is in Section 6. Section 7 concludes. All proofs are given in the appendix, and an online appendix contains supplementary results.

2 The Linear-Rational Framework

We introduce the linear-rational framework and present explicit formulas for zero-coupon bond prices and the short rate. We 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 describe interest rate swaptions, and derive a swaption pricing formula. Finally, the linear-rational framework is compared to 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 $\zeta_t$ 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 has a linear drift, and the state price density is a linear function of the current state. More specifically, we assume that $X_t$ is of the form

$$dX_t = \kappa(\theta - X_t)dt + dM_t$$

(2)

for some $\kappa \in \mathbb{R}^{d \times d}$, $\theta \in \mathbb{R}^d$, and some martingale $M_t$. 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

$$\zeta_t = e^{-\alpha t} (\phi + \psi^T X_t),$$

(3)

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

The linear drift of the factor process implies that conditional expectations are of the following linear form:

$$\mathbb{E}_t[X_T] = \theta + e^{-\kappa(T-t)}(X_t - \theta), \quad t \leq T.$$

(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 why we refer to this framework as linear-rational. Indeed, the basic pricing formula (1) with $C_T = 1$ shows that the zero-coupon bond prices are given by $P(t, T) = F(T - t, X_t)$, where

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

(5)

9One could replace the drift $\kappa(\theta - X_t)$ in (2) with the slightly more general form $b + \beta X_t$ 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 $\beta$) 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.

10This follows from Lemma A.5 in the appendix.
The short rate is obtained via the formula 

\[ r_t = -\frac{\partial T}{\partial t} \log P(t, T)|_{T=t}, \]

and is given by

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

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

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

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

\[ r_t \in [0, \alpha^* - \alpha_*] \quad (r_t \in [0, \infty) \text{ if } \alpha_* = -\infty). \] (8)

Notice that \( \alpha^* \) and \( \alpha_* \) depend on the parameters of the process \( X_t \), which are estimated from data. A crucial step of the model validation process is therefore to verify that the range of possible short rates is sufficiently wide. Finally, whenever the eigenvalues of \( \kappa \) have nonnegative real part, one verifies that \((-1/\tau) \log F(\tau, x)\) converges to \( \alpha \) when \( \tau \) goes to infinity. That is, \( \alpha \) can be interpreted as the infinite-maturity zero-coupon bond yield.

### 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 \( \xi \). It is convenient to carry out this discussion in terms of the kernel of a function.\(^{11}\)

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

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

\(^{11}\)We define the *kernel* of a differentiable function \( f \) on \( E \) by

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

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.
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 \( x \in E \).\(^{12}\)

Therefore the location of the state \( X_t \) along the direction \( \xi \) 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. The following result characterizes \( \mathcal{U} \) in terms of the model parameters.

**Theorem 2.2.** Assume the short rate \( r_t \) is not constant.\(^{13}\) Then \( \mathcal{U} \) is the largest subspace of \( \text{ker} \psi^\top \) that is invariant under \( \kappa \). Formally, this is equivalent to

\[
\mathcal{U} = \text{span} \{ \psi, \kappa^\top \psi, \ldots, \kappa^{(d-1)}^\top \psi \}^\perp.
\]

(9)

In the case where \( \kappa \) is diagonalizable, this leads to the following corollary.

**Corollary 2.3.** Assume \( \kappa \) is diagonalizable with real eigenvalues, i.e. \( \kappa = S^{-1} \Lambda S \) with \( S \) invertible and \( \Lambda \) diagonal and real. Then the term structure kernel is trivial, \( \mathcal{U} = \{0\} \), if and only if all eigenvalues of \( \kappa \) are distinct and all components of \( S^{-\top} \psi \) are nonzero.

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 \( \hat{X}_t = SX_t \) satisfies the linear drift dynamics

\[
d\hat{X}_t = \hat{\kappa}(\hat{\theta} - \hat{X}_t)dt + d\hat{M}_t,
\]

where

\[
\hat{\kappa} = S \kappa S^{-1}, \quad \hat{\theta} = S \theta, \quad \hat{M}_t = SM_t.
\]

(10)

Defining also

\[
\hat{\phi} = \phi, \quad \hat{\psi} = S^{-\top} \psi,
\]

(11)

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

\[
S(\mathcal{U}) = \{0\} \times \mathbb{R}^n
\]

(12)

\(^{12}\)Here and in the sequel, \( \nabla F(\tau, x) \) denotes the gradient with respect to the \( x \) variables.

\(^{13}\)In view of (6), the short rate \( r_t \) is constant if and only if \( \psi \) is an eigenvector of \( \kappa^\top \) with eigenvalue \( \lambda \) 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 (9) equals \( \text{ker} \psi^\top \). The assumption that the short rate is not constant will be in force throughout the paper.
where \( n = \dim U \) and \( m = d - n \). Decomposing the transformed factor process accordingly, \( \tilde{X}_t = (Z_t, U_t) \), 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 (12) holds if and only if the transformed model parameters (10)–(11) satisfy:\(^{14}\)

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

(ii) \( \hat{\kappa} \) has block lower triangular structure, \( \hat{\kappa} = \begin{pmatrix} \hat{\kappa}_{ZZ} & 0 \\ \hat{\kappa}_{UZ} & \hat{\kappa}_{UU} \end{pmatrix} \in \mathbb{R}^{(m+n) \times (m+n)} \);

(iii) The upper left block \( \hat{\kappa}_{ZZ} \) of \( \hat{\kappa} \) satisfies \( \text{span} \{ \hat{\psi}_Z, \hat{\kappa}_{ZZ}^\top \hat{\psi}_Z, \ldots, \hat{\kappa}_{ZZ}^{(m-1)} \hat{\psi}_Z \} = \mathbb{R}^m \).

Assuming that (12) holds, and writing \( S_x = (z, u) \in \mathbb{R}^m \times \mathbb{R}^n \) and \( \hat{\theta} = (\hat{\theta}_Z, \hat{\theta}_U) \), we now see that

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

does not depend on \( u \). Hence, bond prices are given by \( P(t, T) = \hat{F}(T - t, Z_t) \). 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 theorem formalizes the above discussion.

**Theorem 2.5.** The term structure \( \hat{F}(\tau, z) \) is injective if and only if \( \hat{\kappa}_{ZZ} \) is invertible and

\[
\hat{\phi} + \hat{\psi}_Z^\top \hat{\theta}_Z \neq 0. \quad ^{15}
\]

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

\[
dZ_t = \hat{\kappa}_{ZZ} (\hat{\theta}_Z - Z_t) dt + d\hat{M}_Z_t \tag{13}
\]

\(^{14}\)The block forms of \( \hat{\psi} \) and \( \hat{\kappa} \) in (i)–(ii) just reflect that \( \{0\} \times \mathbb{R}^n \) is a subspace of \( \ker \hat{\psi}_Z^\top \) that is invariant under \( \hat{\kappa} \). Condition (iii) asserts that \( \{0\} \times \mathbb{R}^n \) is the largest subspace of \( \ker \hat{\psi}_Z^\top \) with this property.

\(^{15}\)Injectivity here means that if \( \hat{F}(\tau, z) = \hat{F}(\tau, z') \) for all \( \tau \geq 0 \), then \( z = z' \). In other words, if \( \hat{F}(\tau, Z_t) \) is known for all \( \tau \geq 0 \), we can back out the value of \( Z_t \).
and unspanned factor dynamics

\[ dU_t = \left( \tilde{\kappa}_{UZ}(\tilde{\theta}_Z - Z_t) + \tilde{\kappa}_{UU}(\tilde{\theta}_U - U_t) \right) dt + d\tilde{M}_{Ut}, \]  

(14)

where we denote \( \tilde{M}_t = (\tilde{M}_{Zt}, \tilde{M}_{Ut}) \). The state price density can be written

\[ \zeta_t = e^{-\alpha t}(\tilde{\phi} + \tilde{\psi}_Z^\top Z_t). \]  

(15)

The process \( Z_t \) has a linear drift that depends only on \( Z_t \) itself. 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 (13)–(15), which is equivalent to (2)–(3). In view of Theorem 2.2, this leads to an interpretation of Theorem 2.4(iii): the model (13)–(15) is minimal in the sense that its own term structure kernel is trivial. The situation where \( U_t \) enters into the dynamics of \( \tilde{M}_{Zt} \) gives rise to USV. This is discussed in detail for the diffusion case in Section 3.

If our interest is in unspanned factors, why did we not specify the linear-rational model in the \((Z_t, U_t)\)-coordinates in the first place? The reason is that we have to control the interplay of the factor dynamics with the state space. Note that \( E \) has to lie in the half-space \( \{ \phi + \psi^\top x > 0 \} \), and thus always has a non-trivial boundary. The invariance of \( E \) with respect to the factor dynamics \( X_t \) is a non-trivial property, which is much simpler to control if \( E \) has some regular shape (below we will consider \( E = \mathbb{R}^d_+ \)). The shape of the transformed state space \( \tilde{E} = S(E) \) may be deformed, and it is difficult to specify a priori conditions on the \((Z_t, U_t)\)-dynamics (13)–(14) that would assert the invariance of \( \tilde{E} \).

Finally, note that even if the term structure kernel is trivial, \( U = \{0\} \), the short end of the term structure may nonetheless be insensitive to movements of the state along certain directions. In view of Theorem 2.2, for \( d \geq 3 \) we can have \( U = \{0\} \) while still there exists a nonzero vector \( \xi \) such that \( \psi^\top \xi = \psi^\top \kappa \xi = 0 \). This implies that the short rate function is constant along \( \xi \), see (6). On the other hand, we can still, in the generic case, recover \( X_t \) from a snapshot of the term structure, see Theorem 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 $\Delta K$ and the floating leg pays LIBOR accrued over the preceding time period.\footnote{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.} From the perspective of the fixed-rate payer, the value of the swap at time $t \leq T_0$ is given by\footnote{This valuation equation, which was the market standard until a few years ago, implicitly assumes that payments are discounted with an interest 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 Filipović and Trolle (2013). In the present paper we simplify matters by adhering to the formula (16).}

$$\Pi_t^{\text{swap}} = P(t, T_0) - P(t, T_n) - \Delta K \sum_{i=1}^n P(t, T_i).$$ \hspace{1cm} (16)

The time-$t$ forward swap rate, $S_t^{T_0, T_n}$, is the rate $K$ that makes the value of the swap equal to zero. It is given by

$$S_t^{T_0, T_n} = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^n \Delta P(t, T_i)}. \hspace{1cm} (17)$$

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.\footnote{Conversely, a receiver swaption gives the right to enter into an interest rate swap, receiving the fixed leg at a pre-determined rate and paying 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

$$C_{T_0} = (\Pi_{T_0}^{\text{swap}})^+ = \left( \sum_{i=0}^n c_i P(T_0, T_i) \right)^+ = \frac{1}{\zeta_{T_0}} \left( \sum_{i=0}^n c_i E_{T_0} [\zeta_{T_i}] \right)^+, \hspace{1cm}$$

for coefficients $c_i$ that can easily be read off the expression (16).

In a linear-rational term structure model, the conditional expectations $E_{T_0} [\zeta_{T_i}]$ are linear functions of $X_{T_0}$, with coefficients that are explicitly given in terms of the model parameters, see (4). Specifically, we have $C_{T_0} = p_{\text{swap}}(X_{T_0})^+ / \zeta_{T_0}$, where $p_{\text{swap}}$ is the explicit linear function

$$p_{\text{swap}}(x) = \sum_{i=0}^n c_i e^{-\alpha T_i} \left( \phi + \psi^\top \theta + \psi^\top e^{-\kappa (T_i - T_0)} (x - \theta) \right).$$
The swaption price at time $t \leq T_0$ is then obtained by an application of the fundamental pricing formula (1), which yields

$$\Pi^\text{swaption}_t = \frac{1}{\zeta_t} \mathbb{E}_t[\zeta_{T_0} C_{T_0}] = \frac{1}{\zeta_t} \mathbb{E}_t [p_{\text{swap}}(X_{T_0})^+] \quad \text{(18)}.$$

To compute the price, one has to evaluate the conditional expectation on the right side of (18). If the conditional distribution of $X_{T_0}$ is known, this can be done via direct numerical integration over $\mathbb{R}^d$. This is a challenging problem in general; fortunately there is an efficient alternative approach based on Fourier transform methods.

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

$$\Pi^\text{swaption}_t = \frac{1}{\zeta_t} \pi \int_{\mathbb{R}_+} \text{Re} \left[ \hat{q}(\mu + i\lambda) \right] d\lambda.$$

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 $\hat{q}(\mu + i\lambda)$ efficiently as $\lambda$ 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 square-root factor processes, for which computing $\hat{q}(z)$ amounts to solving a system of ordinary differential equations, see Section 4.

It is often more convenient to represent swaption prices in terms of implied volatilities. In the USD market, the market standard is the “normal” implied volatility (NIV), 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.\(^{19}\) When the swaption strike is equal to the forward swap rate ($K = S^{T_0,T_n}_t$, see (17)), there is a particularly simple relation between the swaption price and the NIV, $\sigma_{N,t}$, given by

$$\Pi^\text{swaption}_t = \sqrt{T_0 - t} \cdot \frac{1}{\sqrt{2\pi}} \left( \sum_{i=1}^{n} \Delta P(t,T_i) \right) \sigma_{N,t}; \quad \text{see, e.g., Corp (2012).}$$

\(^{19}\)This is sometimes also referred to as the “absolute” or “basis point” implied volatility. 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.
2.4 Comparison with Other Models

Related to linear-rational term structure models are the Linearity-Generating (LG) processes studied in Gabaix (2009) and Cheridito and Gabaix (2008). In a nutshell, and aligned with our notation, an LG process of dimension $d$ with generator $\kappa \in \mathbb{R}^{d \times d}$ is characterized as $X_t$ along with state price density $\zeta_t$ given as in the linear-rational model (2)–(3) with $\theta = 0$ and $\phi = 0$.\footnote{In Gabaix (2009), the dimension is denoted $d = n + 1$, $X_t$ is denoted $Y_t$, and $\psi$ is denoted $\nu$. The state price density $\zeta_t$ is replaced by the more general expression $M_t D_t$, a pricing kernel $M_t$ times a dividend $D_t$. For zero-coupon bonds we have $D_t = 1$. The factor $e^{-\alpha t}$ does not, strictly speaking, show up in the defining expressions of an LG process in Gabaix (2009). But it is easy to see that the LG process is invariant with respect to this manipulation. Define $Y_t = e^{-\alpha t} X_t$. Then $Y_t$ induces an observationally equivalent LG process $dY_t = -\alpha Y_t dt + e^{-\alpha t} dM_t$ along with $\zeta_t = \psi^\top Y_t$.} Zero-coupon bond prices based on an LG process are strictly linear-rational in $X_t$,

$$P(t, T) = e^{-\alpha (T-t)} \frac{\psi^\top e^{-\kappa (T-t)} X_t}{\psi^\top X_t},$$

or equivalently, strictly linear in the normalized factor $Z_t = X_t / \zeta_t$.\footnote{Cheridito and Gabaix (2008) give sufficient conditions under which the setup with $\zeta_t = \psi^\top X_t$ can be brought via linear transformation $\tilde{X}_t = S X_t$ to $\zeta_t = \psi^\top S^{-1} \tilde{X}_t = \tilde{X}_t$, so that one can assume $\psi = e_1$. That is, the normalized factor satisfies $Z_{1t} = 1$ and $X_t$ has the form $\zeta_t(1, Z_{2t}, \ldots, Z_{dt})$.} Itô calculus shows that the normalized factor process $Z_t$ exhibits quadratic terms in the drift.

An LG process forms a linear-rational model. We now show that this inclusion of model classes is strict. The linear-rational model (2)–(3) is observationally invariant with respect to affine transformations $\tilde{X}_t = S (X_t - b)$ for any invertible linear transformation $S$ and vector $b$ in $\mathbb{R}^d$. Indeed, $\tilde{X}_t$ has linear drift,

$$d\tilde{X}_t = S \kappa S^{-1} \left( S (\theta - b) - \tilde{X}_t \right) dt + S dM_t,$$

and the state price density $\zeta_t = e^{-\alpha t} \left( \phi + \psi^\top b + \psi^\top S^{-1} \tilde{X}_t \right)$ is linear in $\tilde{X}_t$. This yields the following characterization result.

**Lemma 2.7.** The linear-rational model (2)–(3) is observationally equivalent to an LG process of dimension $d$ if and only if $\phi + \psi^\top \theta = \psi^\top \nu$ for some $\nu \in \ker \kappa$. If $\ker \kappa \subseteq \ker \psi^\top$ then this condition reduces to $\phi + \psi^\top \theta = 0$. Hence, term structure models based on LG processes are strictly included in the linear-rational framework.

The specific LG term structure model studied in Carr, Gabaix, and Wu (2009) has a martingale part given by $dM_t = e^{-\alpha t} \beta dN_t$, where $\beta$ is a vector in $\mathbb{R}^d$ and $N_t$ is a scalar expo-
nential martingale of the form $dN_t/N_t = \sum_{i=1}^{m} \sqrt{v_{it}}dB_{it}$ for independent Brownian motions $B_{it}$ and processes $v_{it}$ following square-root dynamics. This LG process is non-stationary due to the time-inhomogeneous volatility specification. The eigenvalues of $\kappa$ have positive real parts, so that the volatility of $X_t$ tends to zero as time goes to infinity, and $X_t$ itself converges to zero almost surely.\footnote{This LG process can be made stationary by multiplying it with $e^{-at}$, as outlined in Footnote 20, if and only if $\beta$ is an eigenvector of $\kappa$ with eigenvalue $\alpha$.} Further, as $N_t$ is scalar, bond prices are perfectly correlated in their model. The linear-rational models we consider in our empirical analysis are time-homogeneous and stationary, and have a volatility structure that is different from the specification in Carr, Gabaix, and Wu (2009), generating a truly multi-factor structure for bond prices.

When the factor process $X_t$ is Markovian, the linear-rational models fall 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 $\zeta_t = e^{-at}R_\alpha g(X_t)$, where $R_\alpha$ 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 $\zeta_t = \int_t^\infty M_{tu} \mu(u) \, du$, 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 $e^{-at}R_\alpha g(X_t) = \int_t^\infty \mathbb{E}_t [e^{-au}g(X_u)] \, du$, which implies $M_{tu} \mu(u) = \mathbb{E}_t [e^{-au}g(X_u)]$. The linear-rational framework fits into this template by taking $\mu(u) = e^{-au}$ and $M_{tu} = \mathbb{E}_t [g(X_u)] = \alpha \phi + \alpha \psi^\top \theta + \psi^\top (\alpha + \kappa)e^{-\kappa(u-t)}(X_t - \theta)$, 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 $\phi$ and $\psi$ positive, $\kappa = \theta = 0$, and letting the martingale part $M_t$ of the factor process $X_t$ be geometric Brownian motion.
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

\[ dX_t = \kappa(\theta - X_t)dt + \sigma(X_t)dB_t \]

(20)

for some \( d \)-dimensional Brownian motion \( B_t \), and some dispersion function \( \sigma(x) \). We denote the diffusion matrix by \( a(x) = \sigma(x)\sigma(x)^\top \), and assume that it is differentiable.

A short calculation using Itô’s formula shows that the dynamics of the state price density can be written

\[ \frac{d\zeta_t}{\zeta_t} = -r_t dt - \lambda_t^\top dB_t \]

where the short rate \( r_t \) is given by (6), and

\[ \lambda_t = -\frac{\sigma(X_t)^\top \psi}{\phi^\top \psi^\top X_t} \]

is the market price of risk. It then follows that the dynamics of \( P(t,T) \) is

\[ \frac{dP(t,T)}{P(t,T)} = (r_t + \nu(t,T)^\top \lambda_t) dt + \nu(t,T)^\top dB_t, \]

(21)

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 the online appendix we confirm this intuition.

We now refine the discussion in Section 2.2 by singling out those cases that give rise to USV. To this end we describe those directions \( \xi \in \mathbb{R}^d \) with the property that movements of the state vector along \( \xi \) do not influence bond return volatilities of any maturity. According to (21), the squared volatility at time \( t \) of the return on the bond with maturity \( T \) is given by \( ||\nu(t,T)||^2 = G(T - t, X_t) \), where we define

\[ G(\tau, x) = \frac{\nabla F(\tau, x)^\top a(x) \nabla F(\tau, x)}{F(\tau, x)^2}. \]

(22)
In analogy with Definition 2.1 we introduce the following notion:

**Definition 3.1.** The volatility kernel, denoted by \( W \), is given by

\[
W = \bigcap_{\tau \geq 0} \ker G(\tau, \cdot).
\]

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

Analogously to Section 2.2 we may transform the state space so that the intersection \( U \cap W \) of the term structure kernel and volatility kernel corresponds to the last components of the state vector. To this end, let \( S \) be an invertible linear transformation satisfying (12), with the additional property that \( S(U \cap W) = \{0\} \times \{0\} \times \mathbb{R}^q \), where \( q = \dim U \cap W \), and \( p + q = n = \dim U \). 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. In contrast, movements of \( V_t \), while having no effect on the term structure, do impact bond return volatilities. 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.\(^{23}\)

Whether the model exhibits USV depends on how the diffusion matrix \( a(x) \) interacts with the other parameters of the model. The following theorem is useful as it gives sufficient conditions for USV, in terms of the coordinates characterized in Theorem 2.4, which can easily be verified in our specifications below.

**Theorem 3.2.** Let \( S \) be any invertible linear transformation satisfying (12), and denote by \( \hat{a}(z, u) = Sa(S^{-1}(z, u))S^\top \) the diffusion matrix of the transformed factor process \( \hat{X}_t = (Z_t, U_t) \). Assume that \( \phi + \psi^\top \theta \neq 0 \) and \( \hat{\kappa}_{ZZ} \) is invertible. Then the number \( p \) of USV factors equals

\[
p = \dim \text{span} \{ \nabla_u \hat{a}_{ij}(z, u) : 1 \leq i \leq j \leq m, (z, u) \in S(E) \}.
\]

\(^{23}\)The “hidden factors \( h_t \)” in Duffee (2011) and the “unspanned components of the macro variables \( M_t \)” in Joslin, Priebsch, and Singleton (2014), both of which are Gaussian exponential-affine models, are residual factors. Indeed, they neither show up in the bond yields nor in the bond volatilities. They enter through the equivalent change of measure from the risk-neutral \( Q \) to the historical measure \( P \), and thus affect the distribution of future bond prices under \( P \), but not under \( Q \). In fact, the volatility kernel \( W \) in a \( d \)-factor Gaussian exponential-affine model is always all of \( \mathbb{R}^d \), since bond return volatilities do not depend on the state.
If \( p = n \) then \( \mathcal{U} \cap \mathcal{W} = \{0\} \), so that every unspanned factor is a USV factor.

4 The Linear-Rational Square-Root Model

The primary example of a linear-rational diffusion model (20) with state space \( E = \mathbb{R}^d_+ \) is the linear-rational square-root (LRSQ) model. It is based on a square-root factor process of the form

\[
dX_t = \kappa(\theta - X_t)dt + \text{Diag} \left( \sigma_1 \sqrt{X_{1t}}, \ldots, \sigma_d \sqrt{X_{dt}} \right) dB_t,
\]

with volatility parameters \( \sigma_i \). In this section we show that USV can easily be incorporated and swaptions can be priced efficiently in the LRSQ model. This lays the groundwork for our empirical analysis.

The aim is 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 (23), we have \( \phi = 1 \) and \( \psi = 1 \), where we write \( 1 = (1, \ldots, 1)^\top \). In this case, the extremal values in (7) are given by \( \alpha^* = \max \mathcal{S} \) and \( \alpha_* = \min \mathcal{S} \), where \( \mathcal{S} = \{1^\top \kappa \theta, -1^\top \kappa_1, \ldots, -1^\top \kappa_d\} \), and \( \kappa_i \) denotes the \( i \)th column vector of \( \kappa \).

In accordance with this result, we always let the state price density be given by \( \zeta_t = e^{-\alpha t}(1 + 1^\top X_t) \) 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 from the observation that (12) holds if and only if the last \( n \) column vectors of \( S^{-1} \) form a basis of the term structure kernel \( \mathcal{U} \). The first \( m \) columns of \( S^{-1} \) can be freely chosen, as long as all column vectors stay linearly independent. The next observation is that \( \ker 1^\top \) is spanned by vectors of the form \(-e_i + e_j\), for \( i < j \), where \( e_i \) denotes the \( i \)th standard basis vector in \( \mathbb{R}^d \). We now choose \(-e_i + e_{m+i}, i = 1, \ldots, n\), as basis for \( \mathcal{U} \), which thus lies in \( \ker 1^\top \) as required by Theorem 2.2. We then choose \( e_1, \ldots, e_m \) as the first \( m \) columns of \( S^{-1} \). This amounts to specifying the invertible linear transformation \( S \) on \( \mathbb{R}^d \) by

\[
S^{-1} = \begin{pmatrix} \text{Id}_m & -A \\ 0 & \text{Id}_n \end{pmatrix} \quad \text{with} \quad A \in \mathbb{R}^{m \times n} \text{ defined by } A = \begin{pmatrix} \text{Id}_n \\ 0 \end{pmatrix}
\]
so that

\[ S = \begin{pmatrix} \text{Id}_m & A \\ 0 & \text{Id}_n \end{pmatrix}. \]

The parameters appearing in the description (23) of the factor process \( X_t \) can now be specified with the aid of Theorem 2.4. First, by construction we have that \( \hat{\psi} = S^{-\top}1 = (1, 0) \) is the vector in \( \mathbb{R}^d \) whose first \( m \) components are ones and the remaining components are zeros. This confirms Theorem 2.4(i). For convenience we now introduce the index sets \( I = \{1, \ldots, m\} \) and \( J = \{m + 1, \ldots, d\} \). We write the mean reversion matrix \( \kappa \) in block form as

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

where \( \kappa_{IJ} \) denotes the submatrix whose rows are indexed by \( I \) and columns by \( J \), and similarly for \( \kappa_{II}, \kappa_{JI}, \kappa_{JJ} \). The transformed mean reversion matrix

\[ \hat{\kappa} = S\kappa S^{-1} = \begin{pmatrix} \kappa_{II} + A\kappa_{JI} & -\kappa_{II}A - A\kappa_{JI} + \kappa_{IJ} + A\kappa_{JJ} \\ \kappa_{JI} & \kappa_{JJ} - \kappa_{JI}A \end{pmatrix} \]

becomes block-triangular if and only if \( \kappa_{IJ} = \kappa_{II}A - A\kappa_{JI} + A\kappa_{JJ} \), which complies with Theorem 2.4(ii). For the sake of parsimony we also assume that \( \kappa_{JI} = 0 \) and \( \kappa_{JJ} = A^\top\kappa_{II}A \), which is the upper left \( n \times n \) block of \( \kappa_{II} \), so that \( \hat{\kappa} \) becomes block-diagonal. Theorem 2.4(iii) then holds if and only if

\[ \text{span} \left\{ 1, \kappa_{II}^\top 1, \ldots, \kappa_{II}^{(m-1)} \right\} = \mathbb{R}^m. \tag{24} \]

This construction motivates the following definition.

**Definition 4.2.** The LRSQ(m,n) specification is obtained by choosing \( \kappa_{II} \in \mathbb{R}^{m \times m} \) with nonpositive off-diagonal elements and such that (24) holds. The mean reversion matrix is defined by

\[ \kappa = \begin{pmatrix} \kappa_{II} & \kappa_{II}A - AA^\top\kappa_{II}A \\ 0 & A^\top\kappa_{II}A \end{pmatrix}. \]

The level of mean reversion is taken to be a vector \( \theta \in \mathbb{R}^d \) with \( \kappa\theta \in \mathbb{R}_+^d \), and the volatility parameters are taken to be nonnegative, \( \sigma_1, \ldots, \sigma_d \geq 0 \).

This definition guarantees that a unique solution to (23) exists, see e.g. Filipović (2009, Theorem 10.2). Indeed, note that \( \kappa \) has nonpositive off-diagonal elements by construction.
The following theorem confirms that the $LRSQ(m,n)$ specification indeed exhibits USV.

**Theorem 4.3.** The $LRSQ(m,n)$ specification exhibits $m$ term structure factors and $n$ unspanned factors. Assume that $1 + 1^T \theta \neq 0$ and $\kappa_{11}$ is invertible. Then the number of USV factors equals the number of indices $1 \leq i \leq n$ such that $\sigma_i \neq \sigma_{m+i}$. If $\sigma_i \neq \sigma_{m+i}$ for all $1 \leq i \leq n$ then every unspanned factor is a USV factor.

As an illustration, we 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.$^{24}$

**Example 4.4.** In 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 $\kappa$, $\widehat{\kappa} = \kappa$, and the corresponding volatility matrix is

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

Theorem 4.3 implies that $U_t$ is a USV factor if $\sigma_1 \neq \sigma_2$, $\kappa_{11} \neq 0$, and $1 + 1^T \theta \neq 0$.

Swaption pricing becomes particularly tractable in the LRSQ model. Since $X_t$ is a square-root process, 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 well-known system of ordinary differential equations; see, e.g., Duffie, Pan, and Singleton (2000) and Filipović (2009, Theorem 10.3). For convenience the relevant expressions are reproduced in the online appendix. In order for Theorem 2.6 to be applicable, it is necessary that some exponential moments of $p_{\text{swap}}(X_{T_0})$ be finite. We therefore note that for any $v \in \mathbb{R}^d$ there is always some $\mu > 0$ (depending on $v$, $X_0$, $T_0$) such that $\mathbb{E}[\exp(\mu v^T X_{T_0})] < \infty$. While it may be difficult a priori to decide how small $\mu$ should be, the choice is easy in practice since numerical methods diverge if $\mu$ is too large, resulting in easily detectable outliers.$^{24}$

---

$^{24}$This contradicts the statement of Collin-Dufresne and Goldstein (2002, Proposition 3) that a two-factor Markov model of the term structure cannot exhibit USV.
5 Flexible Market Price of Risk Specification

The market price of risk in a linear-rational diffusion model is endogenous, and can be too restrictive to match certain empirical features of the data. In this section we describe a simple way of introducing flexibility in the market price of risk specification, which allows us to circumvent this issue.25

The starting point is the observation that the linear-rational framework can equally well be developed under some auxiliary probability measure \( \mathbb{A} \) that is equivalent to the historical probability measure \( \mathbb{P} \) with Radon–Nikodym density process \( \mathbb{E}_t [d\mathbb{A}/d\mathbb{P}] \). The state price density \( \zeta_t \equiv \zeta_t^\mathbb{A} = e^{-\alpha t}(\phi + \psi^\top X_t) \) and the martingale \( M_t \equiv M_t^\mathbb{A} \) are then understood with respect to \( \mathbb{A} \). The factor process dynamics reads

\[
dX_t = \kappa(\theta - X_t)dt + dM_t^\mathbb{A},
\]

and the basic pricing formula (1) becomes \( \Pi(t,T) = \mathbb{E}_t^\mathbb{A} [\zeta_T^\mathbb{A} C_T] / \zeta_t^\mathbb{A} \). From this, using Bayes’ rule, we obtain the state price density with respect to \( \mathbb{P} \), \( \zeta_t^\mathbb{P} = \zeta_t^\mathbb{A} \mathbb{E}_t [d\mathbb{A}/d\mathbb{P}] \), so that \( \Pi(t,T) = \mathbb{E}_t^\mathbb{P} [\zeta_T^\mathbb{P} C_T] / \zeta_t^\mathbb{P} \). Bond prices \( P(t,T) = F(T-t, X_t) \) are still given as functions of the factor process \( X_t \), with the same \( F(\tau, x) \) as in (5).

In the diffusion setup of Section 3, the martingale \( M_t^\mathbb{A} \) is now given by \( dM_t^\mathbb{A} = \sigma(X_t)dB_t^\mathbb{A} \) for some \( \mathbb{A} \)-Brownian motion \( B_t^\mathbb{A} \), and the market price of risk \( \lambda_t \equiv \lambda_t^\mathbb{A} = -\sigma(X_t)^\top \psi/(\phi + \psi^\top X_t) \) is understood with respect to \( \mathbb{A} \). Having specified the model under the auxiliary measure \( \mathbb{A} \), we now have full freedom in choosing an equivalent change of measure from \( \mathbb{A} \) to the historical measure \( \mathbb{P} \). Specifically, \( \mathbb{P} \) can be defined using a Radon–Nikodym density process of the form

\[
\mathbb{E}_t \left[ \frac{d\mathbb{P}}{d\mathbb{A}} \right] = \exp \left( \int_0^t \delta_s^\top dB_s^\mathbb{A} - \frac{1}{2} \int_0^t \|\delta_s\|^2 ds \right),
\]

for some appropriate integrand \( \delta_t \). The \( \mathbb{P} \)-dynamics of the factor process becomes

\[
dX_t = (\kappa(\theta - X_t) + \sigma(X_t)\delta_t) dt + \sigma(X_t)dB_t^\mathbb{P},
\]

for the \( \mathbb{P} \)-Brownian motion \( dB_t^\mathbb{P} = dB_t^\mathbb{A} - \delta_t dt \). The state price density with respect to \( \mathbb{P} \) follows the dynamics \( d\zeta_t^\mathbb{P}/\zeta_t^\mathbb{P} = -r_t dt - (\lambda_t^\mathbb{P})^\top dB_t^\mathbb{P} \), where the market price of risk \( \lambda_t^\mathbb{P} \) is now

25The same idea has been used in Rogers (1997).
given by

\[ \lambda_t^P = \lambda_t^B + \delta_t = -\frac{\sigma(X_t)^T \psi}{\phi + \psi^T X_t} + \delta_t. \]

The exogenous choice of \( \delta_t \) gives us the freedom to introduce additional unspanned factors. Bond return volatility vectors are invariant under an equivalent change of measure. Hence, the squared volatility of the return on the bond with maturity \( T \), \( \|\nu(t,T)\|^2 = G(T - t, X_t) \), is still given as function of the factor process, with the same \( G(\tau, x) \) as in (22). As a consequence, unspanned factors entering through \( \delta_t \) only are residual factors, affecting risk premiums but not bond return volatilities.\(^{26}\) In our empirical analysis the focus is on USV factors, and our specification of \( \delta_t \) in Section 6.2 below does not introduce additional unspanned factors to the model.

6 Empirical Analysis

We perform an extensive empirical analysis of the LRSQ model. We focus in particular on how the model captures key features of term structure, volatility, and risk premium dynamics.

6.1 Data

The empirical analysis is based 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 1, 2, 3, 5, 7, and 10 years, respectively. We also observe prices of swaptions with the same six swap maturities, option expiries of 3 months and 1, 2, and 5 years, and strikes equal to the forward swap rates. Such at-the-money-forward swaptions are the most liquid. We convert swaption prices into NIVs using (19) with zero-coupon bond prices bootstrapped from the swap curve. The data is from Bloomberg and consists of composite quotes based on quotes from major banks and inter-dealer brokers. Each time series consists of 866 weekly observations from January 29, 1997 to August 28, 2013.

Time series of the 1-year, 5-year, and 10-year swap rates are displayed in Panel A1 of Figure 2. The 1-year swap rate fluctuates between a minimum of 0.30% (on May 1, 2013) and a maximum of 7.51% (on May 17, 2000), while the longer-term swap rates exhibit less

\(^{26}\)This is the type of unspanned factors in Duffee (2011) and Joslin, Priebsch, and Singleton (2014), see Footnote 23.
variation. A principal component analysis (PCA) of weekly changes in swap rates shows that the first three factors explain 90%, 7%, and 2%, respectively, of the variation. Panel B1 of Figure 2 displays time series of NIVs of three “benchmark” swaptions: the 3-month option on the 2-year swap, the 2-year option on the 2-year swap, and the 5-year option on the 5-year swap. Of these, the 3-month NIV of the 2-year swap rate is the most volatile, fluctuating between a minimum of 18 bps (on December 12, 2012) and a maximum of 213 bps (on October 8, 2008). Swaptions also display a high degree of commonality, with the first three factors from a PCA of weekly changes in NIVs explaining 77%, 8%, and 5%, respectively, of the variation. Summary statistics of the data are given in the online appendix.

6.2 Model Specifications

We restrict attention to the LRSQ(m,n) specification, see Definition 4.2. We always set \( m = 3 \) (three term structure factors) 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). The online appendix provides the explicit factor dynamics in these specifications.

We develop the model under the \( A \)-measure and obtain the \( P \)-dynamics of the factor process by specifying \( \delta_t \) parsimoniously as

\[
\delta_t = \left( \sqrt{X_{1t}}, \ldots, \sqrt{X_{dt}} \right)^\top.
\]

This choice is convenient as \( X_t \) remains a square-root process under \( P \), facilitating the use of standard estimation techniques from the vast body of literature on affine models (note, however, that \( X_t \) is not a square-root process under \( Q \)).\(^{27}\) Specifically, we estimate by quasi-maximum likelihood in conjunction with Kalman filtering.\(^{28}\) Details are provided in the online appendix.

In preliminary analyses, we find that the upper-triangular elements of \( \kappa_{II} \) are always

\(^{27}\)Alternatively, one could use a measure change from \( A \) to \( P \) similar to the one suggested by Cheridito, Filipovic, and Kimmel (2007). In this case, \( X_t \) would also be a square-root process under \( P \), but the model would be less parsimonious.

\(^{28}\)With Kalman filtering, all swaps and swaptions are subject to measurement/pricing errors. Alternatively, one could assume that particular portfolios of swaps and swaptions are observed without error and invert the factors from those portfolios. Due to the nonlinear relation between swaps/swaptions and the factors, this approach is not practical in our setting. In the context of estimating Gaussian term structure models with zero-coupon bond yields, Joslin, Singleton, and Zhu (2011) show that the two approaches give very similar results.
very close to zero. The same is true of the lower left element of $\kappa_{II}$. As a first step towards obtaining more parsimonious specifications, we reestimate the models after setting these elements to zero; i.e., imposing that $\kappa_{II}$, and thus $\hat{\kappa}_{ZZ}$, is lower bi-diagonal. Furthermore, several of the parameters in (25) are very imprecisely estimated. As a second step towards obtaining more parsimonious specifications, we follow Duffee (2002) and Dai and Singleton (2002) in reestimating the models after setting to zero those market price of risk parameters for which the absolute $t$-statistics did not exceed one. In all cases, the likelihood functions are virtually unaffected by these constraints.

6.3 Maximum Likelihood Estimates

Table 1 displays parameter estimates and their asymptotic standard errors. It is straightforward to verify that (24) holds true in all model specifications implying that the number of term structure factors cannot be reduced to less than three. Also, a robust feature across all specifications is that the drift parameters align such that to a close approximation

$$\alpha = 1^T \hat{\kappa}_{ZZ} \hat{\theta}_Z = -1^T \hat{\kappa}_{ZZ,1} = -1^T \hat{\kappa}_{ZZ,2} > -1^T \hat{\kappa}_{ZZ,3} = \alpha_*,$$

where $\hat{\kappa}_{ZZ,i}$ denotes the $i$th column vector of $\hat{\kappa}_{ZZ}$. Combined with $1^T \hat{\kappa}_{ZZ,3} = \kappa_{33}$, the range of the short rate (8) is given by $r_t \in [0, \alpha + \kappa_{33}]$, and the expression (6) for $r_t$ effectively reduces to

$$r_t = (\alpha + \kappa_{33}) \frac{Z_{3t}}{1 + 1^T Z_t}. \quad (26)$$

It is immediately clear from this expression that the ZLB on the short rate is attained when $Z_{3t} = 0$. The table reports the upper bound on the short rate, which is 20.0%, 146.1%, and 71.8% for the $LRSQ(3,1)$, $LRSQ(3,2)$, and $LRSQ(3,3)$ specification, respectively. As such, the upper bound is not a restrictive feature of the model. Simulations show that the likelihood of observing very high short rates is negligible in all specifications; in contrast, as shown below there is a significant likelihood of observing very low short rates. $\alpha$ ranges between 5.66% and 7.46% across model specifications, which appears reasonable given that $\alpha$ has the economic interpretation as the model-implied infinite-maturity zero-coupon bond

---

29 The historical maximum for the effective federal funds rate is 22.4%. However, such levels of short rates were only reached during the monetary policy experiment (MPE) in the early 1980s when the Federal Reserve was conducting monetary policy in a fundamentally different way by targeting monetary aggregates instead of interest rates. Regime-switching models such as Gray (1996) and Dai, Singleton, and Yang (2007) also identify the period around the MPE as a very different regime than the post-MPE period.
yield (all specifications are stationary since all eigenvalues of $\hat{\kappa}_{ZZ}$ are positive).\textsuperscript{30}

From a practical perspective, the simple structure for the short rate gives the model flexibility to capture significant variation in longer-term interest rates during the latter part of the sample period when policy rates are effectively zero. The reason is that a low value of $Z_{3t}$ constrains the short rate to be close to zero, which allows $Z_{1t}$ and $Z_{2t}$ the freedom to affect longer-term interest rates without having much impact on the short rate.

Finally, note that the conditions for every unspanned factor to be a USV factor are met; see Theorem 4.3. Specifically, for all $1 \leq i \leq n$, we have $\sigma_{i+3} > \sigma_i$. Since the instantaneous volatility of the $i$’th term structure factor is given by $\sqrt{\sigma_i^2 Z_{it} + (\sigma_{i+3}^2 - \sigma_i^2) U_{it}}$, term structure factor volatilities are increasing in the USV factors.

### 6.4 Factors

Figure 3 displays the estimated factors. The first, second, and third column show the factors of the $LRSQ(3,1)$, $LRSQ(3,2)$, and $LRSQ(3,3)$ model specification, respectively. The first, second, and third row correspond to $(Z_{1t}, U_{1t})$, $(Z_{2t}, U_{2t})$, and $(Z_{3t}, U_{3t})$, respectively. The factors are highly correlated across specifications, which is indicative of a stable factor structure. The USV factors are occasionally large relative to the term structure factors, which gives the first indication of the importance of allowing for USV.

In terms of the economic interpretation of the term structure factors, it follows from (26) that $Z_{3t}$ is highly correlated with the short end of the yield curve. Taking the $LRSQ(3,3)$ specification as an example, the correlation between $Z_{3t}$ and the 1-year swap rate is 0.84 in weekly changes (0.98 in levels). For the other two term structure factors, $Z_{1t}$ is most highly correlated with the slope of the term structure (in weekly changes, the correlation with the spread between the 10-year and 1-year swap rates is 0.65), while $Z_{2t}$ is most highly correlated with the curvature of the term structure (in weekly changes, the correlation with the 1-5-10 “butterfly” spread—twice the 5-year rate swap minus the sum of the 1-year and 10-year swap rates—is 0.72).

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 show results for the $LRSQ(3,1)$, $LRSQ(3,2)$, and $LRSQ(3,3)$ specification, respectively, while the first, second, and third row correspond to $Z_{1t}$, $Z_{2t}$, and $Z_{3t}$, respectively. The grey areas

\textsuperscript{30}As an out-of-sample check on the value of $\alpha$, we bootstrapped the swap curve out to 30 years on each observation date (recall that only swap maturities up to 10 years are used in the estimation). The sample mean of the 30-year zero-coupon bond yield is 5.26%.
mark the possible range of factor volatilities, which is given by $\sigma_i \sqrt{Z_{it}}$ to $\sigma_{i+3} \sqrt{Z_{it}}$ in case that the $i$’th term structure factor exhibits USV. Whenever USV is allowed for, there appears to be significant variation in factor volatilities that is unrelated to the factor levels.

Finally, we consider the correlation between the term structure factors. In the model, the factors are instantaneously uncorrelated, while the unconditional correlations may be non-negative due to the feedback via the drift. However, within the branch of exponential-affine term structure models for which the factor process is also of the square-root type (the $A_m(m)$ model in the notation of Dai and Singleton (2000)), the unconditional correlations among the estimated factors are often reported to be strongly negative (see, e.g., the discussion in Dai and Singleton (2000) of the Duffie and Singleton (1997) model). In our setting, this appears not to be the case. Taking again the $LRSQ(3,3)$ specification as an example, the unconditional correlation is strongly positive between $Z_{2t}$ and $Z_{3t}$ (0.82), somewhat positive between $Z_{1t}$ and $Z_{2t}$ (0.19), and virtually zero between $Z_{1t}$ and $Z_{3t}$ (0.01). This is consistent with the lower bi-diagonal structure of $\hat{\kappa}_{ZZ}$.

### 6.5 Specification Analysis

For each of the model specifications, we compute the fitted swap rates and NIVs based on the filtered state variables. We then compute weekly root-mean-squared pricing errors (RMSEs) for swap rates and RMSEs for NIVs—both across all swaptions and for each option expiry separately. The first three rows in Table 2 report the sample means of the RMSE time series for the three model specifications. The next two rows report the mean differences in RMSEs between model specifications along with the associated Newey and West (1987) $t$-statistics in parentheses.

The most parsimonious $LRSQ(3,1)$ specification has a reasonable fit to the data, with mean RMSEs for swap rates and NIVs equal to 7.11 bps and 6.63 bps, respectively. Adding one more USV factor decreases the mean RMSEs by 3.28 bps and 0.86 bps, respectively, which is both economically important and statistically significant ($t$-statistics of -8.95 and -2.18, respectively). The improvement in the fit to swaptions mainly occurs at the 2-year and 5-year option expiries. Adding an additional USV factor has a negligible impact on the mean RMSE for swap rates, but decreases the mean RMSE for NIVs by a further 0.58 bps, which again is both economically important and statistically significant ($t$-statistic of -2.52). In this case, the improvement in the fit to swaptions mainly occurs at the 3-month and 1-year option expiries.
The performance of the $LRSQ(3,3)$ specification over time is illustrated in Figure 2 which shows time series of fitted swap rates (Panel A2) and NIVs (Panel B2) as well as time series of RMSEs for swap rates (Panel A3) and NIVs (Panel B3). The RMSEs are relatively stable over time, except during the financial crisis when the RMSEs are generally larger. The transitory spikes in RMSEs are mostly associated with familiar crisis events, such as the collapse of Long-Term Capital Management in September 1998, the bond market sell-off in June 2003 driven by MBS convexity hedging, and the default of Lehman Brothers in October 2008.

To investigate the performance of the model when policy rates are close to the ZLB, we also split the sample period into a ZLB period and a pre-ZLB period. The beginning of the ZLB period 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. Taking the $LRSQ(3,3)$ specification as an example, the mean RMSEs for swap rates and NIVs are 2.97 bps and 4.78 bps, respectively, during the pre-ZLB period compared with 5.59 bps and 6.21 bps, respectively, during the ZLB period. Given the challenging market conditions during the ZLB period, including potentially distortive effects from the Federal Reserve’s quantitative easing programs, this performance seems respectable. The online appendix contains additional information on the fit of the model specifications.

6.6 Term Structure Dynamics near the Zero Lower Bound

A key characteristic of the recent history of U.S. interest rates is the extended period of near-zero policy rates. Even more strikingly, Japan has experienced near-zero policy rates since the early 2000s. A challenge for term structure models is to generate such extended periods of low short rates. A related challenge, as emphasized by Kim and Singleton (2012), is that near the ZLB, the distribution of future short rates becomes highly asymmetric with the most likely (modal) values being significantly lower than the mean values.

To evaluate the model along this dimension, we simulate 50,000 years of weekly data (2,600,000 observations) from each of the three model specifications.\footnote{The processes are simulated using the “full truncation” Euler discretization scheme of Lord, Koekkoek, and van Dijk (2010). To minimize the discretization bias, we use 10 steps per day.} We use simulated data instead of fitted data because the simulated data is much more revealing about the true properties of the model (this also applies to the analyses in the subsequent sections). Taking the $LRSQ(3,3)$ specification as an example and conditioning on the short rate being between
0 and 25 bps, Panels A-C in Figure 5 display histograms showing the frequency distributions of the future short rate at a 1-year, 2-year, and 5-year horizon, respectively. Clearly the model generates persistently low short rates; conditional on the short rate being between 0 and 25 bps at a given point in time, the likelihood of the short rate being in the same interval after 1, 2, and 5 years is 64%, 47%, and 24%, respectively. Also, the conditional distributions of the future short rates are highly asymmetric; the most likely interval for the short rate remains 0 to 25 bps at all displayed horizons, while Panel D shows the mean (median) value of the short rate rising to 1.12% (0.80%) at a 5-year horizon.\textsuperscript{32}

The ability to generate persistently low short rates is related to the simple structure for the dynamics of the term structure factors in conjunction with the effective expression for $r_t$ in (26). Simple computations show that $Z_{1t}$ is the only term structure factor with a non-negligible constant term in its drift function. This implies that the mean-reversion level of $Z_{3t}$ under $\mathbb{P}$ is determined by $Z_{2t}$ (and possibly $U_{3t}$), while that of $Z_{2t}$ is determined by $Z_{1t}$ (and possibly $U_{2t}$). Therefore, starting from $Z_{3t} \approx 0$, $Z_{3t}$ (and $r_t$) does not necessarily drift immediately away from zero; rather the drift of $Z_{3t}$ (and $r_t$) can be constrained by a low value of $Z_{2t}$, the drift of which can be constrained by a low value of $Z_{1t}$. It appears that this “cascading” structure causes $r_t$ to only slowly drift away from zero.

Another consequence of extended periods of low short rates is a changing characteristic of the factor loadings, particularly the “level” factor. Figure 6 shows the factor loadings of the first principal component of the term structure when the short rate is away from the ZLB (black lines) and close to the ZLB (grey lines). The factor loadings are constructed from the covariance matrix of weekly changes in swap rates. Solid lines correspond to the data. During the pre-ZLB period, the loadings are relatively flat (in fact somewhat hump-shaped) justifying its “level” factor label. However, during the ZLB period, the loadings increase strongly with maturity rising from 0.13 for the 1-year maturity to 0.52 for the 10-year maturity.\textsuperscript{33} As such, the first factor effectively becomes a “slope” factor.

The figure also shows the factor loadings in the simulated weekly data, in which the ZLB sample consists of those observations where the short rate is less than 25 bps. Dashed,

\textsuperscript{32}The online appendix shows that the $LRSQ(3,2)$ specification generates similar conditional distributions, while the $LRSQ(3,1)$ specification generates somewhat more persistently low short rates. Kim and Singleton (2012) show that quadratic models and shadow-rate models also have the ability to generate persistently low short rates. Note that they focus on the conditional $\mathbb{Q}$-distribution, while the focus here is on the conditional $\mathbb{P}$-distribution.

\textsuperscript{33}In contemporaneous work, Kim and Priebsch (2013) also note the changing characteristics of factor loadings during the ZLB period.
dashed-dotted, and dotted lines correspond to the \(LRSQ(3,1)\), \(LRSQ(3,2)\), and \(LRSQ(3,3)\) specifications, respectively. In general, the model-implied factor loadings are close to those observed in the data, particular for the ZLB sample. As such, the model captures how the level factor morphs into a slope factor when the short rate is near the ZLB.

### 6.7 Volatility Dynamics near 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. Using data that pre-dates the current environment of very low interest rates, several papers find that a large component of volatility is only weakly related to the term structure.\(^{34}\) Here, we provide an important qualification to this result: volatility becomes compressed and gradually more level-dependent as interest rates approach the ZLB. This is illustrated by Figure 1, which, as mentioned in the Introduction, shows the 3-month NIV of the 1-year swap rate (in bps) plotted against the level of the 1-year swap rate.\(^{35}\) The grey area in Figure 1 marks the possible range of implied volatilities in case of the \(LRSQ(3,3)\) specification and shows that the model qualitatively captures the observed pattern.

More formally, for each swap maturity, we regress weekly changes in the 3-month NIV of the swap rate on weekly changes in the level of the swap rate (including a constant); i.e.,

\[
\Delta \sigma_{N,t} = \beta_0 + \beta_1 \Delta S_t + \epsilon_t.
\]

We run these regressions unconditionally as well as conditional on swap rates being in the intervals 0%-1%, 1%-2%, 2%-3%, 3%-4%, and 4%-5%, respectively.\(^{36}\) Result are displayed in Table 3 in which Panel A shows \(\hat{\beta}_1\)s with Newey and West (1987) \(t\)-statistics in parentheses, and Panel B shows \(R^2\)s. Within each panel, the first row displays results of the unconditional regressions, while the second to sixth rows display results of the conditional regressions. The right-most column reports average \(\hat{\beta}_1\)s and \(R^2\)s across swap maturities. In the unconditional regressions, the \(\hat{\beta}_1\)s are positive but relatively small in magnitudes (between 0.16 and 0.18),

---

\(^{34}\)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.

\(^{35}\)In the online appendix, we overlay data from the four largest swap markets—the U.S., the Eurozone, the U.K, and Japan—and show that a similar pattern is observed for all the swap maturities considered in the paper (we include international data to increase the number of data points with very low interest rates).

\(^{36}\)We run the regressions in first-differences to avoid spurious results due to the persistence of implied volatilities and interest rates.
and, though the coefficients are statistically significant ($t$-statistics between 2.80 and 5.49), the $R^2$s are small (between 0.05 and 0.11). That is, unconditionally, there is a relatively small degree of positive level-dependence in volatility.\footnote{37}

A more nuanced picture emerges from the conditional regressions. Conditional on swap rates being in the interval 0%-1%, the $\hat{\beta}_1$s are positive, large in magnitudes (between 0.48 and 1.20), and very highly statistically significant ($t$-statistics between 7.83 and 8.79 despite fewer observations than in the unconditional regressions), and the $R^2$s are large (between 0.44 and 0.54). In other words, there is a strong and positive relation between volatility and swap rate changes when swap rates are close to the ZLB.\footnote{38} However, as the conditioning interval increases in level, the relation between volatility and swap rate changes becomes progressively weaker, and volatility exhibits very little level-dependence at moderate levels of swap rates. For instance, conditional on swap rates being in the interval 4%-5%, the $\hat{\beta}_1$s are close to zero (between -0.07 and 0.08) and mostly statistically insignificant ($t$-statistics between -0.82 and 1.76), and the $R^2$s are very small (between 0.00 and 0.03).

We then perform the same analysis on the simulated weekly data. To succinctly summarize the results, we focus on the $LRSQ(3,3)$ specification and its ability to capture the average $\hat{\beta}_1$s and $R^2$s across swap maturities. Figure 7 shows the average $\hat{\beta}_1$s and $R^2$s in the data (Panels A and C) and those implied by the model (Panels B and D). In the unconditional regressions, the average $\hat{\beta}_1$ and $R^2$ in the data are 0.16 and 0.08, respectively, while the corresponding model-implied values are 0.15 and 0.15, respectively. In the conditional regressions, the average $\hat{\beta}_1$ and $R^2$ in the data are 0.76 and 0.51, respectively, when swap rates are between 0% and 1%, decreasing to 0.03 and 0.02, respectively, when swap rates are between 4% and 5%. This is closely matched by the model, which generates values of 0.72 and 0.52, respectively, when swap rates are between 0% and 1%, decreasing to 0.10 and 0.12, respectively, when swap rates are between 4% and 5%. As such, the model largely matches the changing volatility dynamics as interest rates approach the ZLB.\footnote{39}

\footnote{37}Trolle and Schwartz (2014) also document that, unconditionally, there is a positive level-dependence in NIVs. 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 monetary policy experiment in the early 1980s, which is not representative of the current monetary policy regime.\footnote{38}In Japanese data, Kim and Singleton (2012) also note the high degree of level-dependence in volatility when interest rates are close to the ZLB.\footnote{39}The online appendix shows that the $LRSQ(3,2)$ specification generates similar volatility dynamics, while the $LRSQ(3,1)$ specification generates too high a degree of level-dependence in volatility, both unconditionally and conditionally, as a consequence of only having one USV factor.\footnote{29}
6.8 Risk Premium Dynamics

Next, we study swap risk premiums. We first establish key properties of swap risk premiums and then investigate if the model is able to capture those properties. Returns on swap contracts can be computed in several ways. We work with excess returns on forward-starting swaps, which are most easily computed from the available swap data. The online appendix shows that very similar results are obtained with excess returns on spot-starting swaps as well as zero-coupon bonds bootstrapped from swap rates.

Consider a holding period of $\Delta$ and the swap contract described in Section 2.3 that lasts from $T_0$ to $T_n$. The strategy enters into the forward-starting swap contract at time $t = T_0 - \Delta$, receiving fixed and paying floating, and allocates an amount of capital $C_t$ (covering at least the required margins) earning the risk-free rate. At time $T_0$, the value of the swap is

$$\Pi_{T_0}^{\text{swap}} = -(1 - P(T_0, T_n)) + S_{T_0}^{T_0, T_n} \left( \sum_{i=1}^{n} \Delta P(T_0, T_i) \right) = -(S_{T_0}^{T_0, T_n} - S_{T_0}^{T_0, T_n}) \left( \sum_{i=1}^{n} \Delta P(T_0, T_i) \right).$$

Since the swap has zero initial value, the holding period excess return is

$$R_{T_0}^{T} = \frac{\Pi_{T_0}^{\text{swap}}}{C_t}.$$ 

We consider nonoverlapping monthly excess returns (i.e., $\Delta = 1$ month) computed using closing prices on the last business day of each month (we have verified that our results hold true regardless of the day of the month that the strategy is initiated). Furthermore, we assume that $C_t$ equals the notional of the swap (equal to one) corresponding to a “fully collateralized” swap position.

Table 4 shows the unconditional means and volatilities of excess returns in addition to the unconditional Sharpe ratios. All statistics are annualized. Both the mean and volatility increase with swap maturity, while the Sharpe ratio decreases with swap maturity. A similar pattern for Sharpe ratios has been observed by Duffee (2010) and Frazzini and Pedersen (2014) in the case of Treasury bonds. The mean excess returns and Sharpe ratios are high in our sample, but are inflated by the downward trend in swap rates over the sample period.

We also consider conditional expected excess returns on swap contracts. For each swap maturity, we regress nonoverlapping monthly excess swap returns on previous month’s term

40Alternatively, one could choose $C_t$ to achieve a certain target volatility for excess returns as in Duarte, Longstaff, and Yu (2007).
structure slope and implied volatility (including a constant); i.e.,

\[ R_{t,T_0}^c = \beta_0 + \beta_{Slp}Slp_t + \beta_{Vol}Vol_t + \epsilon_{t,T_0}, \]

where \( Slp_t \) is the difference between the spot swap rate with a \((T_n - T_0)\)-year maturity and 1-month LIBOR, and \( Vol_t \) is the NIV of a swaption with a 1-month option expiry and a \((T_n - T_0)\)-year swap maturity. We use a 1-month option expiry to match the horizon in the predictive regression. Excess returns are in percent, and both \( Slp_t \) and \( Vol_t \) are standardized to facilitate comparison of regression coefficients. Results are reported in Table 5 with Newey and West (1987) \( t \)-statistics given in parentheses. Many papers, typically using long samples of Treasury data, find that the slope of the term structure predicts excess bond returns with a positive sign; see, e.g., Campbell and Shiller (1991) and Dai and Singleton (2002). In our more recent (and shorter) sample of swap data, the predictive power of the slope of the term structure is relatively weak. Indeed, the regression coefficient is never statistically significant and is even negative for 1-year and 2-year swap maturities. In contrast, the predictive power of implied volatility is much stronger. The regression coefficient is statistically significant for swap maturities up to 3 years and is positive for all swap maturities indicating a positive risk-return tradeoff in the swap market. This economically, volatility is also a more important predictor of excess returns. Taking the 5-year swap maturity as an example, a one standard deviation increase in volatility (the term structure slope) increases the one-month expected excess returns by 16 bps (8 bps) which should be put in relation to the unconditional mean of the one-month excess return of 28 bps.

To investigate if our model is able to capture these characteristics of risk premiums, we simulate 50,000 years of monthly data (600,000 observations) from each of the three model specifications. Note that swaptions with 1-month option expiries are not included in the estimation, giving an out-of-sample flavor to the exercise. The results are displayed in the lower panels in Tables 4 and 5. Table 4 shows that all model specifications capture the

\[ 41 \text{For the Treasury market, there is mixed evidence for a risk-return tradeoff. Engle, Lilien, and Robins (1987) and, more recently, Ghysels, Le, Park, and Zhu (2014) find evidence of a positive risk-return tradeoff in an ARCH-in-mean framework, while Duffee (2002) runs regressions similar to those in Table 5 and finds that volatility only weakly predicts return. The differences between the results of Duffee (2002) and our results are likely due to some combination of differences in sample periods (his sample ends where our sample begins), his use of historical volatilities vs. our use of forward-looking implied volatilities, and structural differences between the Treasury and swap markets. Note also that many equilibrium term structure models including those within the long-run-risk framework (such as Bansal and Shaliastovich (2012)) and habit-based framework (such as Wachter (2006)) generally imply a positive risk-return tradeoff; see also the discussion in Le and Singleton (2013).} \]
pattern that the mean and volatility of excess returns increase with swap maturity, while the Sharpe ratio decreases with swap maturity (there is a small hump in the Sharpe ratio term structure for the $LRSQ(3,3)$ specification). The mean excess returns and Sharpe ratios are lower than those observed in the data, which is to be expected given that we are simulating stationary samples of swap rates, while the swap rates in the data exhibit a downward trend over the sample period.

Table 5 shows that the model qualitatively captures the predictive power of implied volatility for excess returns. The size of the regression coefficients as well as the $R^2$s increase with the number of USV factors. For the $LRSQ(3,3)$ specification, the ratio between model-implied and actual $\hat{\beta}_{vol}$ ($R^2$) lie between 0.45 and 0.77 (0.67 and 1.00) across swap maturities. This performance appears reasonable considering the parsimonious market price of risk specification that we employ.\footnote{As a further “sanity check” of the estimates of the market prices of risk, we follow Duffee (2010) in considering the model-implied maximum conditional Sharpe ratio, which is given by $\sqrt{(\lambda P_t)\top(\lambda P_t)}$ in our setting. Based on the simulated data, we infer that the population means of this quantity are 1.40, 0.91, and 0.64 for the $LRSQ(3,1)$, $LRSQ(3,2)$, and $LRSQ(3,3)$ specifications, respectively. For comparison, Duffee (2010) reports that for a standard three-factor Gaussian term structure model, and considering log-returns, the population mean is 1.07 (when annualized). As such, our market price of risk estimates appear reasonable.}

### 6.9 Comparison with Exponential-Affine Model

It is instructive to compare the LRSQ model with the exponential-affine model for which the factor process is also of the square-root type (the $A_m(m)$ model). The two models share the property that interest rates are nonnegative, but the latter model is severely restricted by not being able to accommodate USV and not admitting analytical solutions to swaptions. To put the two models on an equal footing, we contrast the $LRSQ(m,0)$ model with the $A_m(m)$ model and focus solely on their abilities to price swaps.

In the $A_m(m)$ model, the short rate is a linear function of the current state, $r_t = \gamma \top X_t$ for some $\gamma \in \mathbb{R}^m_+$, and the $Q$-dynamics of the factor process is given by

$$dX_t = \kappa(\hat{\theta} - X_t)dt + \text{Diag} \left(\sqrt{X_{1t}}, \ldots, \sqrt{X_{mt}}\right) dB^Q_t, \quad (27)$$

where $\kappa$ and $\hat{\theta}$ satisfy standard admissibility conditions and $B^Q_t$ denotes a $Q$-Brownian motion.\footnote{We follow Kim and Singleton (2012) in normalizing the diffusion parameters in (27) to ones (to achieve identification) and setting the constant term in the short-rate expression to zero (to impose a lower bound of zero on the short rate as we do for the LRSQ model).} Zero-coupon bond prices are exponential-affine—instead of linear-rational—functions.
of the current state with \( P(t,T) = e^{A(T-t)+B(T-t)^T X_t} \), where \( A(\tau) \) and \( B(\tau) \) solve a well-known system of ordinary differential equations. The \( \mathbb{P} \)-dynamics of the factor process is obtained by setting the market price of risk equal to \( \delta_t \) given in (25). This aligns the two models as closely as possible—they have the same \( \mathbb{P} \)-dynamics of the factor processes and the same number of model parameters—making it easier to isolate their structural differences.

We estimate both models with \( m = 3 \) using the swap data described in Section 6.1 and again applying quasi-maximum likelihood in conjunction with Kalman filtering. Parameter estimates and other details are given in the online appendix. The two models differ in the set of parameters that can be identified from the term structure, which is \( (\tilde{\kappa}, \tilde{\theta}, \gamma) \) in the exponential-affine model compared with \( (\kappa, \theta) \) in the linear-rational model. However, despite the linear-rational model having a more parsimonious description of the term structure, the pricing performance of the two models is virtually identical both for the entire sample period (mean RMSE of 2.73 bps for \( LRSQ(3,0) \) vs. 2.72 bps for \( A_3(3) \)) and for the ZLB period (mean RMSE of 4.48 bps for \( LRSQ(3,0) \) vs. 4.53 bps for \( A_3(3) \)). In the linear-rational model, the diffusion parameters are only identified from the time-series of the factors, which makes the model-implied factor volatilities more in line with actual ones. To see this, we first estimated actual factor volatilities by fitting EGARCH models to the innovations of the filtered factors. We then tested whether the mean differences between the model-implied and actual factor volatilities are statistically significant. Based on Newey and West (1987) standard errors with long lags, the mean difference is statistically significant for two of the factors in the \( A_3(3) \) model, but insignificant for all of the factors in the \( LRSQ(3,0) \) model.\(^{44}\)

In the linear-rational model, the drift (under both \( \mathbb{P} \) and \( \mathbb{Q} \) ) and instantaneous variance of the short rate are nonlinear functions of the factors—in contrast to the exponential-affine model, where these functions are linear. Specifically, the drift functions of \( r_t \) under \( \mathbb{P} \) and \( \mathbb{Q} \) are given by

\[
\nabla r(X_t)^\top \left(\kappa(\theta - X_t) + \text{Diag}(\sigma_1\delta_1, \ldots, \sigma_3\delta_3) X_t - \frac{a(X_t)1}{1+1^\top X_t}\right) \quad \text{and} \quad \nabla r(X_t)^\top \kappa(\theta - X_t),
\]

respectively, and the instantaneous variance function is given by

\[
\nabla r(X_t)^\top a(X_t)\nabla r(X_t),
\]

\(^{44}\)We only consider whether the models capture the factor volatilities on average. Given the absence of USV in these models, much of the variation in factor volatilities is not captured by the models.
where $\nabla r(x) = (\kappa^\top (1 + 1^\top x) + 1^\top \kappa(\theta - x)) 1/(1 + 1^\top x)^2$, and $a(x) = \text{Diag}(\sigma_1^2 x_1, \ldots, \sigma_3^2 x_3)$ is the diffusion matrix. To compare the two models in terms of their short-rate dynamics, we first simulate 50,000 years of weekly observations of the state variables. We then set the state vector equal to its mean value and compute the drift and instantaneous variance as we vary each of the factors between its minimum and maximum value attained in the simulation. Panels A1 and A2 in Figure 8 display the drift under $\mathbb{P}$ (solid lines) and $\mathbb{Q}$ (dash-dotted lines) for the $LRSQ(3,0)$ and $A_3(3)$ models, respectively. Panels B1 and B2 display the instantaneous variance. In the figure, the factors are normalized so that they range between zero and one. It appears that the linear-rational model exhibits a moderate degree of nonlinearity in both the drift and instantaneous variance.

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 nonnegative interest rates, ii) easily accommodates unspanned factors affecting volatility and risk premiums, and iii) admits analytical solutions to swaptions. A parsimonious model specification within the linear-rational class has a very good fit to both interest rate swaps and swaptions from 1997 to 2013 and captures many features of term structure, volatility, and risk premium dynamics—including when interest rates are close to the ZLB.

Many extensions are possible. The $LRSQ(m,n)$ specification is constructed to be as parsimonious as possible. Following the recipe in Section 4, more intricate specifications can be constructed. Also, we focus on unspanned factors that affect volatility but also show up as factors affecting risk premiums. As discussed in Section 5, it is possible to introduce pure unspanned risk premium factors in the sense of Duffee (2011) and Joslin, Priebsch, and Singleton (2014). More generally, the linear-rational framework can be extended or modified to be applicable to equity, credit, and commodity markets.

\footnote{In the $A_3(3)$ model, the first two elements of $\gamma$ are essentially zero; therefore, the instantaneous variance of $r_t$ only depends on $X_3$. Duarte (2004) (in continuous time) and Le, Singleton, and Dai (2010) (in discrete time) develop exponential-affine term structure models which are nonlinear under $\mathbb{P}$ (but not under $\mathbb{Q}$). We leave a comparison with those models for future research.}
A Proofs

This appendix provides the proofs of all theorems and corollaries in the paper. Proofs of the lemmas are deferred to the online appendix.

Proof of Theorem 2.2

By the Cayley-Hamilton theorem (see Horn and Johnson (1990, Theorem 2.4.2)) we know that any power $\kappa^p$ can be written as linear combination of $\text{Id}$, $\kappa^T$, $\ldots$, $\kappa^{(d-1)}^T$. Taking orthogonal complements in (9), we thus need to prove

$$\text{span}\left\{\nabla F(\tau, x) : \tau \geq 0, x \in E \right\} = \text{span}\left\{\kappa^p \psi : p \geq 0 \right\}. \quad (28)$$

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

$$\nabla F(\tau, x) = \frac{e^{-\alpha \tau}}{\phi + \psi^T} x \left[ e^{-\kappa^T \tau} \psi - e^{\alpha \tau} F(\tau, x) \psi \right], \quad (29)$$

whence $S = \text{span}\{e^{-\kappa^T \tau} \psi - e^{\alpha \tau} F(\tau, x) \psi : \tau \geq 0, x \in E\}$. By the assumption that the short rate is not constant, there are $x, y \in E$ and $\tau \geq 0$ such that $F(\tau, x) \neq F(\tau, y)$. It follows that $e^{\alpha \tau} (F(\tau, x) - F(\tau, y)) \psi$, and hence $\psi$ itself, lies in $S$. We deduce that $S = \text{span}\{e^{-\kappa^T \tau} \psi : \tau \geq 0\}$, which coincides with the right side of (28). This proves the formal expression (9).

It remains to show that $U$ given by (9) equals $U'$, defined as the largest subspace of ker $\psi^T$ that is invariant under $\kappa$. It follows from (9), for $p = 0$, that $U$ is a subspace of ker $\psi^T$. Moreover, $U$ is invariant under $\kappa$: let $\xi \in U$. Then $\kappa \xi \in \ker \psi^T \kappa^{p-1}$ for all $p \geq 1$, and hence $\kappa \xi \in U$. Since $U'$ is the largest subspace of $\psi^T$ with this property, we conclude that $U \subseteq U'$. Conversely, let $\xi \in U'$. By invariance we have $\kappa^p \xi \in U' \subseteq \ker \psi^T$, and hence $\xi \in \ker \psi^T \kappa^p$, for any $p \geq 0$. Hence $\xi \in U$, and we conclude that $U' \subseteq U$. 

35
Proof of Corollary 2.3

Write $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_d)$ and consider the matrix $A = [\psi \kappa^T \psi \cdots \kappa^{(d-1)}^T \psi]$. Writing $\hat{\psi} = S^{-T} \psi$, the determinant of $A$ is given by

$$
\det A = \det (S^T) \det \left( \hat{\psi} \Lambda \hat{\psi} \cdots \Lambda^{d-1} \hat{\psi} \right)
= \det (S^T) \hat{\psi}_1 \cdots \hat{\psi}_d \det \begin{pmatrix}
1 & \lambda_1 & \cdots & \lambda_1^{d-1} \\
: & : & & : \\
1 & \lambda_d & \cdots & \lambda_d^{d-1}
\end{pmatrix}
= \det (S^T) \hat{\psi}_1 \cdots \hat{\psi}_d \prod_{1 \leq i < j \leq d} (\lambda_j - \lambda_i),
$$

where the last equality uses the formula for the determinant of the Vandermonde matrix. Theorem 2.2 now shows that the term structure kernel is trivial, $\mathcal{U} = \{0\}$, precisely when all eigenvalues of $\kappa$ are distinct and all components of $\hat{\psi}$ are nonzero, as was to be shown.

Proof of Theorem 2.4

The proof of the sufficiency direction of Theorem 2.4 will build on the following lemma. It provides a useful condition that guarantees the existence of a minimal number of unspanned factors.

Lemma A.1. Let $m, n \geq 0$ be integers with $m + n = d$. If the transformed model parameters (10)–(11) satisfy (i)–(ii) in Theorem 2.4, we have $S(\mathcal{U}) \supseteq \{0\} \times \mathbb{R}^n$. In this case, we have $\dim \mathcal{U} \geq n$.

We can now proceed to the proof of Theorem 2.4. The case $n = 0$ is immediate, so we consider the case $n \geq 1$. We write $\hat{\mathcal{U}} = S(\mathcal{U})$ and assume (12) holds. That is, we have

$$
\hat{\mathcal{U}} = \{0\} \times \mathbb{R}^n \subset \mathbb{R}^m \times \mathbb{R}^n. \quad (30)
$$

On the other hand, Theorem 2.2 yields

$$
\hat{\mathcal{U}} = \left\{ \hat{\xi} \in \mathbb{R}^d : \hat{\psi}^T \hat{\kappa}^p \hat{\xi} = 0 \text{ for } p = 0, 1, \ldots, d-1 \right\}. \quad (31)
$$

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

$$
\hat{\psi} = \begin{pmatrix}
\hat{\psi}_Z \\
\hat{\psi}_U
\end{pmatrix} \in \mathbb{R}^{m+n}, \quad \hat{\kappa} = \begin{pmatrix}
\hat{\kappa}_{ZZ} & \hat{\kappa}_{ZU} \\
\hat{\kappa}_{UZ} & \hat{\kappa}_{UU}
\end{pmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}.
$$
It then follows from (30) and (31) that we have \( \hat{\psi}_U \hat{\xi}_U = 0 \) for any \( \hat{\xi}_U \in \mathbb{R}^n \). Hence \( \hat{\psi}_U = 0 \), which proves (i). As a consequence, we have \( \hat{\psi}_{U} = 0 \), which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]

which proves (i). As a consequence, we have 

\[
\hat{\psi}_{U} = 0,
\]
(ii) \( \kappa \) is invertible and \( \mathcal{U} = \{0\} \);

(iii) \( \psi \in \text{span} \{ \kappa^T \psi, \ldots, \kappa^d \psi \} \) and \( \mathcal{U} = \{0\} \).

We can now proceed to the proof of Theorem 2.5. We first assume \( \mathcal{U} = \{0\} \), so that \( m = d \). Let \( x, y \in E \). In view of the power series expansion of \( F(\tau, x) \) in \( \tau \),

\[
F(\tau, x) = e^{-\alpha \tau} \left( 1 + \sum_{\rho \geq 1} \frac{\psi^T \kappa^\rho (x - \theta)}{\phi + \psi^T x} (-\tau)^\rho \right),
\]

we have \( F(\tau, x) = F(\tau, y) \) for all \( \tau \geq 0 \) if and only if

\[
\frac{\psi^T \kappa^\rho (x - \theta)}{\phi + \psi^T x} = \frac{\psi^T \kappa^\rho (y - \theta)}{\phi + \psi^T y}, \quad \rho \geq 1.
\]

To prove sufficiency, assume \( \kappa \) is invertible and \( \phi + \psi^T \theta \neq 0 \). Then pick \( x, y \) such that (35) is satisfied; we must prove that \( x = y \). Lemma A.2(ii)–(i) implies that we may find coefficients \( a_1, \ldots, a_d \) so that \( \psi = \sum_{\rho=1}^{d} a_p \kappa^\rho \psi \). Multiplying both sides of (35) by \( a_p \) and summing over \( \rho = 1, \ldots, d \) yields

\[
\frac{\psi^T (x - \theta)}{\phi + \psi^T x} = \frac{\psi^T (y - \theta)}{\phi + \psi^T y},
\]

or, equivalently, \( \psi^T (x - y)(\phi + \psi^T \theta) = 0 \). Since \( \phi + \psi^T \theta \neq 0 \) we then deduce from (35) that \( \psi^T \kappa^\rho (x - y) = 0 \) for all \( \rho \geq 0 \), which by the aforementioned spanning property of \( \psi^T \kappa^\rho \), \( \rho \geq 0 \), 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 \( \kappa \) is invertible and \( \phi + \psi^T \theta \neq 0 \). There are two cases. First, assume \( \kappa \) is not invertible. We claim that there is an element \( \eta \in \text{ker} \kappa \) such that \( \theta + s \eta \) lies in the set \( \{ x \in \mathbb{R}^d : \phi + \psi^T x \neq 0 \} \) for all large \( s \). Indeed, if this were not the case we would have \( \text{ker} \kappa \subseteq \text{ker} \psi^T \), which would contradict \( \mathcal{U} = \{0\} \). So such an \( \eta \) exists. Now simply take \( x = \theta + s_1 \eta \), \( y = \theta + s_2 \eta \) for large enough \( s_1 \neq s_2 \)—clearly (35) holds for this choice, proving that injectivity fails.

The second case is where \( \kappa \) is invertible, but \( \phi + \psi^T \theta = 0 \). In particular \( \phi + \psi^T x = \psi^T (x - \theta) \). Together with the fact that \( \kappa^T \psi, \ldots, \kappa^d \psi \) span \( \mathbb{R}^d \) (Lemma A.2(ii)–(i)), this shows that (35) is equivalent to

\[
\frac{x - \theta}{\psi^T (x - \theta)} = \frac{y - \theta}{\psi^T (y - \theta)}.
\]

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 theorem is now complete for the case $U = \{0\}$. The case where $U \neq \{0\}$ follows by applying the above case to the model with factor process $Z$ and state price density $\zeta_t = e^{-at}(\hat{\phi} + \hat{\psi}^T_Z Z_t)$. Indeed, this model has a trivial term structure kernel in view of Theorem 2.4 (iii).

**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))):

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

Let $q(ds)$ denote the conditional distribution of the random variable $p_{\text{swap}}(X_{T_0})$ so that

$$\hat{q}(z) = \int_{\mathbb{R}} e^{zs} q(ds)$$

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

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

where the second equality follows from Tonelli’s theorem. This justifies applying Fubini’s theorem in the following calculation, which uses the identity (36) on the second line:

$$\mathbb{E}_t \left[p_{\text{swap}}(X_{T_0})^+\right] = \int_{\mathbb{R}} s^+ q(ds) = \int_{\mathbb{R}} \left( \frac{1}{2\pi} \int_{\mathbb{R}} e^{(\mu+i\lambda)s} \frac{1}{(\mu+i\lambda)^2} d\lambda \right) q(ds)$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\hat{q}(\mu+i\lambda)}{(\mu+i\lambda)^2} d\lambda = \frac{1}{\pi} \int_0^\infty \Re \left[ \frac{\hat{q}(\mu+i\lambda)}{(\mu+i\lambda)^2} \right] d\lambda.$$

Here the last equality uses that the left, and hence right, side is real, together with the observation that the real part of $(\mu+i\lambda)^{-2}\hat{q}(\mu+i\lambda)$ is an even function of $\lambda$ (this follows from a brief calculation.) The resulting expression for the conditional expectation, together with (18), gives the result.
Proof of Theorem 3.2

Whether a given linear-rational term structure model exhibits USV depends on how the diffusion matrix function $a(x)$ interacts with the other parameters of the model, as Lemma A.4 below shows. Its proof builds on the following lemma.

Lemma A.3. Assume $\phi + \psi^T \theta \neq 0$, and consider any $x \in E$. The following conditions are equivalent.

(i) $\psi \in \text{span} \{ \kappa^T \psi, \ldots, \kappa^d T \psi \}$,

(ii) $U^\perp = \text{span} \{ \nabla F(\tau, x) : \tau \geq 0 \}$.

Lemma A.4. The volatility kernel satisfies

$$U \cap W \supseteq U \cap \bigcap_{\eta_1, \eta_2 \in U^\perp} \ker \eta_1^T a(\cdot) \eta_2$$

with equality if $\phi + \psi^T \theta \neq 0$ and $\psi \in \text{span} \{ \kappa^T \psi, \ldots, \kappa^d T \psi \}$.

We can re-express Lemma A.4 in the coordinates characterized in Theorem 2.4. First note the identity $\nabla_x \eta_1^T a(\hat{x}) \eta_2 = S^T \hat{\nabla} \hat{\eta}_1^T \hat{a}(\hat{x}) \hat{\eta}_2$ for $\hat{x} = Sx$, where $\hat{\eta}_i = S^{-T} \eta_i$. Using the fact that $\eta \in U^\perp$ if and only if $\hat{\eta} \in S(U)^\perp = \mathbb{R}^m \times \{0_n\}$, Lemma A.4 implies that

$$S(U \cap W) \supseteq S(U) \cap \bigcap_{\eta_1, \eta_2 \in U^\perp} S \ker \eta_1^T a(\cdot) \eta_2$$

$$= (\{0_m\} \times \mathbb{R}^n) \cap \bigcap_{1 \leq i \leq j \leq m} \left\{ \hat{\xi} \in \mathbb{R}^d : \nabla_{z} \hat{a}_{ij}(\hat{x})^T \hat{\xi} = 0 \text{ for all } \hat{x} \in S(E) \right\}$$

$$= \{0_m\} \times \bigcap_{1 \leq i \leq j \leq m} \left\{ \hat{\xi}_U \in \mathbb{R}^n : \nabla_{u} \hat{a}_{ij}(z, u)^T \hat{\xi}_U = 0 \text{ for all } (z, u) \in S(E) \right\}.$$}

with equality if $\phi + \psi^T \theta \neq 0$ and $\psi \in \text{span} \{ \kappa^T \psi, \ldots, \kappa^d T \psi \}$. We claim that $\psi \in \text{span} \{ \kappa^T \psi, \ldots, \kappa^d T \psi \}$ if and only if $\hat{\kappa}_{ZZ}$ is invertible. This finishes the proof of Theorem 3.2.

It remains to prove the claim. Indeed, $\psi \in \text{span} \{ \kappa^T \psi, \ldots, \kappa^d T \psi \}$ is equivalent to $\hat{\psi}_Z \in \text{span} \{ \hat{\kappa}^T_{ZZ} \hat{\psi}_Z, \ldots, \hat{\kappa}^m_{ZZ} \hat{\psi}_Z \}$, which follows from Theorem 2.4(i) and (ii). Theorem 2.4(iii) and Lemma A.2(iii)–(ii) now yield the claim.

Proof of Theorem 4.1

The following lemma is used in the proof of Lemma A.6 below.
Lemma A.5. Assume that $X$ is of the form (2) with integrable starting point $X_0$. Then for any bounded stopping time $\rho$ and any deterministic $\tau \geq 0$, the random variable $X_{\rho + \tau}$ is integrable, and we have
\[
\mathbb{E}_\rho[X_{\rho + \tau}] = \theta + e^{-\kappa \tau} (X_\rho - \theta).
\]

We next state a result, which is valid for linear-rational term structure models (2)–(3) with a semimartingale factor process $X_t$ whose minimal state space is the nonnegative orthant $\mathbb{R}^d_+$. Hereby, we say that the state space $E$ is minimal if $\mathbb{P}(X_t \in U \text{ for some } t \geq 0) > 0$ holds for any relatively open subset $U \subset E$.

Lemma A.6. Assume $X_t$ is a semimartingale of the form (2) whose minimal state space is $\mathbb{R}^d_+$. Then $\kappa_{ij} \leq 0$ for all $i \neq j$.

Now consider a linear-rational term structure model (2)–(3) with a semimartingale factor process $X_t$ and minimal state space $\mathbb{R}^d_+$. Since $\phi + \psi^\top x$ is assumed positive on $\mathbb{R}^d_+$, we must have $\psi \in \mathbb{R}^d_+$ and $\phi > 0$. Dividing $\zeta_t$ by $\phi$ does not affect any model prices, so we may take $\phi = 1$. Moreover, after permuting and scaling the components, $X_t$ is still a semimartingale of the form (2) with minimal state space $\mathbb{R}^d_+$, so we can assume $\psi = 1_p$. Here, we let $1_p$ denote the vector in $\mathbb{R}^d$ whose first $p$ ($p \leq d$) components are ones, and the remaining components are zeros. As before, we write $1 = 1_d$. The short rate is then given by $r_t = \alpha - \rho(X_t)$, where
\[
\rho(x) = \frac{1_p^\top \kappa \theta - 1_p^\top \kappa x}{1 + 1_p^\top x} = \frac{1_p^\top \kappa \theta + \sum_{i=1}^d (-1_p^\top \kappa_i) x_i}{1 + \sum_{i=1}^p x_i} \quad (37)
\]
and where $\kappa_i$ denotes the $i$th column vector of $\kappa$.

Lemma A.7. The short rates are bounded from below, $\alpha^* = \sup_{x \in \mathbb{R}^d_+} \rho(x) < \infty$, if and only if $1_p^\top \kappa_i = 0$ for $i > p$. In this case, $\alpha^* = \max S_p$ and $\alpha_* = \min S_p$, where $S_p = \{1_p^\top \kappa \theta, -1_p^\top \kappa_1, \ldots, -1_p^\top \kappa_p\}$, and the submatrix $\kappa_{1\ldots p,p+1\ldots d}$ is zero.

We can now prove Theorem 4.1. To this end, we first observe that the factor process $X_t$ remains a square-root process after coordinatewise scaling and permutation of its components. Hence, as above, we can assume that $\phi = 1$ and $\psi = 1_p$ for some $p \leq d$. Lemma A.7 then shows that short rates are bounded from below if and only if submatrix $\kappa_{1\ldots p,p+1\ldots d}$ vanishes. If $p = d$ there is nothing left to prove. So assume now that $p < d$. This implies that $(X_{1t}, \ldots, X_{pt})$ is an autonomous square-root process on the smaller state space $\mathbb{R}^p_+$. Since
the state price density $\zeta_t$ 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 always take $p = d$, as desired.

Finally, the expressions for $\alpha^*$ and $\alpha_*$ follow directly from Lemma A.7. This completes the proof of Theorem 4.1.

**Proof of Theorem 4.3**

Conditions Theorem 2.4(i)–(iii) are satisfied by construction. Hence by that theorem the $LRSQ(m,n)$ specification exhibits $m$ term structure factors and $n$ unspanned factors. A calculation shows that

$$\hat{a}_{ij}(z,u) = \begin{cases} 
\sigma^2_i z_i + \left(\sigma^2_{m+i} - \sigma^2_i\right) u_i, & \text{if } 1 \leq i = j \leq n, \\
\sigma^2_i z_i, & \text{if } n + 1 \leq i = j \leq m, \\
0, & \text{otherwise}.
\end{cases}$$

This implies

$$\nabla_u \hat{a}_{ij}(z,u) = \begin{cases} 
\left(\sigma^2_{m+i} - \sigma^2_i\right) e_i, & \text{if } 1 \leq i = j \leq n, \\
0, & \text{otherwise}
\end{cases}$$

where $e_i$ denotes the $i$th standard basis vector in $\mathbb{R}^n$. The assertion about the number of USV factors now follows from Theorem 3.2.
<table>
<thead>
<tr>
<th></th>
<th>$\kappa_{11}$</th>
<th>$\kappa_{22}$</th>
<th>$\kappa_{33}$</th>
<th>$\kappa_{21}$</th>
<th>$\kappa_{32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LRSQ(3, 1)$</td>
<td>0.0736</td>
<td>0.4233</td>
<td>0.1258</td>
<td>-0.1482</td>
<td>-0.4979</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0008)</td>
<td>(0.0004)</td>
<td>(0.0003)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>$LRSQ(3, 2)$</td>
<td>0.2910</td>
<td>0.0690</td>
<td>1.3925</td>
<td>-0.3598</td>
<td>-0.1378</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0002)</td>
<td>(0.0044)</td>
<td>(0.0007)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$LRSQ(3, 3)$</td>
<td>0.1895</td>
<td>0.1280</td>
<td>0.6616</td>
<td>-0.2460</td>
<td>-0.1846</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0003)</td>
<td>(0.0018)</td>
<td>(0.0006)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>$\theta_1$</td>
<td>$\theta_2$</td>
<td>$\theta_3$</td>
<td>$\theta_4$</td>
<td>$\theta_5$</td>
</tr>
<tr>
<td>$LRSQ(3, 1)$</td>
<td>0.5455</td>
<td>0.3547</td>
<td>1.4044</td>
<td>0.4677</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td>(0.0012)</td>
<td>(0.0049)</td>
<td>(0.0094)</td>
<td></td>
</tr>
<tr>
<td>$LRSQ(3, 2)$</td>
<td>0.2239</td>
<td>1.1669</td>
<td>0.1220</td>
<td>0.0125</td>
<td>0.0653</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0055)</td>
<td>(0.0006)</td>
<td>(0.0013)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>$LRSQ(3, 3)$</td>
<td>0.2350</td>
<td>0.4517</td>
<td>0.1260</td>
<td>0.0635</td>
<td>0.1221</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0018)</td>
<td>(0.0004)</td>
<td>(0.0005)</td>
<td>(0.0008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_1$</td>
<td>$\sigma_2$</td>
<td>$\sigma_3$</td>
<td>$\sigma_4$</td>
<td>$\sigma_5$</td>
</tr>
<tr>
<td>$LRSQ(3, 1)$</td>
<td>0.3109</td>
<td>0.7052</td>
<td>0.1172</td>
<td>1.6696</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0016)</td>
<td>(0.0010)</td>
<td>(0.0060)</td>
<td></td>
</tr>
<tr>
<td>$LRSQ(3, 2)$</td>
<td>0.5073</td>
<td>0.2749</td>
<td>0.0498</td>
<td>0.6461</td>
<td>1.1917</td>
</tr>
<tr>
<td></td>
<td>(0.0021)</td>
<td>(0.0008)</td>
<td>(0.0004)</td>
<td>(0.0022)</td>
<td>(0.0059)</td>
</tr>
<tr>
<td>$LRSQ(3, 3)$</td>
<td>0.3177</td>
<td>0.2676</td>
<td>0.0509</td>
<td>0.7920</td>
<td>0.9135</td>
</tr>
<tr>
<td></td>
<td>(0.0025)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
<td>(0.0020)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td></td>
<td>$\delta_1$</td>
<td>$\delta_2$</td>
<td>$\delta_3$</td>
<td>$\delta_4$</td>
<td>$\delta_5$</td>
</tr>
<tr>
<td>$LRSQ(3, 1)$</td>
<td>0</td>
<td>0.1178</td>
<td>-2.5141</td>
<td>-0.0431</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1009)</td>
<td>(1.0823)</td>
<td>(0.0354)</td>
<td></td>
</tr>
<tr>
<td>$LRSQ(3, 2)$</td>
<td>0</td>
<td>-0.3074</td>
<td>-5.1463</td>
<td>0.4186</td>
<td>-0.6276</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2748)</td>
<td>(2.2028)</td>
<td>(0.4018)</td>
<td>(0.3509)</td>
</tr>
<tr>
<td>$LRSQ(3, 3)$</td>
<td>0</td>
<td>-0.5095</td>
<td>-0.6592</td>
<td>0.1130</td>
<td>-0.1872</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3188)</td>
<td>(0.6048)</td>
<td>(0.0976)</td>
<td>(0.0526)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\text{sup}_{r_t}$</td>
<td>$\mathcal{L} \times 10^{-4}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$LRSQ(3, 1)$</td>
<td>$\sigma_{\text{rates}}$</td>
<td>$\sigma_{\text{swaptions}}$</td>
<td>0.0746</td>
<td>0.2004</td>
<td>14.6913</td>
</tr>
<tr>
<td></td>
<td>8.3978</td>
<td>7.1879</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0558)</td>
<td>(0.0196)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$LRSQ(3, 2)$</td>
<td>$\delta_1$</td>
<td>$\delta_2$</td>
<td>$\delta_3$</td>
<td>$\delta_4$</td>
<td>$\delta_5$</td>
</tr>
<tr>
<td>$LRSQ(3, 2)$</td>
<td>$\sigma_{\text{rates}}$</td>
<td>$\sigma_{\text{swaptions}}$</td>
<td>0.0688</td>
<td>1.4613</td>
<td>15.0869</td>
</tr>
<tr>
<td></td>
<td>5.0389</td>
<td>6.5017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0393)</td>
<td>(0.0131)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$LRSQ(3, 3)$</td>
<td>$\sigma_{\text{rates}}$</td>
<td>$\sigma_{\text{swaptions}}$</td>
<td>0.0566</td>
<td>0.7181</td>
<td>15.3100</td>
</tr>
<tr>
<td></td>
<td>4.8086</td>
<td>5.6982</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0487)</td>
<td>(0.0137)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Maximum likelihood estimates
The table reports parameter likelihood estimates with asymptotic standard errors are in parentheses. $\sigma_{\text{rates}}$ denotes the standard deviation of swap rate pricing errors and $\sigma_{\text{swaptions}}$ denotes the standard deviation of swaption pricing errors in terms of normal implied volatilities. Both $\sigma_{\text{rates}}$ and $\sigma_{\text{swaptions}}$ are measured in basis points. $\alpha$ is chosen as the smallest value that guarantees a nonnegative short rate. $\text{sup}_{r_t}$ is the upper bound on possible short rates. $\mathcal{L}$ denotes the log-likelihood value. The sample period consists of 866 weekly observations from January 29, 1997 to August 28, 2013.
Table 2: Comparison of model specifications.
The table reports means of time series of the root-mean-squared pricing errors (RMSEs) of swap rates and normal implied swaption volatilities. For swaptions, results are reported for the entire volatility surface as well as for the volatility term structures at the four option maturities in the sample (3 months, 1 year, 2 years, and 5 years). Units are basis points. t-statistics, corrected for heteroscedasticity and serial correlation up to 26 lags (i.e. 6 months) using the method of Newey and West (1987), are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The sample period consists of 866 weekly observations from January 29, 1997 to August 28, 2013.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Swaps</th>
<th>Swaptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>3 mths</td>
</tr>
<tr>
<td>$LRSQ(3,1)$</td>
<td>7.11</td>
<td>6.63</td>
</tr>
<tr>
<td>$LRSQ(3,2)$</td>
<td>3.83</td>
<td>5.77</td>
</tr>
<tr>
<td>$LRSQ(3,3)$</td>
<td>3.72</td>
<td>5.19</td>
</tr>
<tr>
<td>$LRSQ(3,2)-LRSQ(3,1)$</td>
<td>$-3.28^{***}$</td>
<td>$-0.86^{**}$</td>
</tr>
<tr>
<td></td>
<td>(-8.95)</td>
<td>(-2.18)</td>
</tr>
<tr>
<td>$LRSQ(3,3)-LRSQ(3,2)$</td>
<td>$-0.12$</td>
<td>$-0.58^{**}$</td>
</tr>
<tr>
<td></td>
<td>(-0.78)</td>
<td>(-2.52)</td>
</tr>
</tbody>
</table>
Table 3: Level-dependence in volatility.
For each available swap maturity, the table reports results from regressing weekly changes in the 3-month normal implied volatility of the swap rate on weekly changes in the level of the swap rate (including a constant). Panel A shows the slope coefficients with \( t \)-statistics in parentheses, and Panel B shows the \( R^2 \)s. Within each panel, the first row displays unconditional results, while the second to sixth rows display results conditional on swap rates being in the intervals 0%-1%, 1%-2%, 2%-3%, 3%-4%, and 4%-5%, respectively. Each underlying time series consists of 866 weekly observations from January 29, 1997 to August 28, 2013. \( t \)-statistics are corrected for heteroscedasticity and serial correlation up to 12 lags using the method of Newey and West (1987). *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

<table>
<thead>
<tr>
<th>Panel A: ( \beta_1 )</th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.18***</td>
<td>0.16***</td>
<td>0.16***</td>
<td>0.16***</td>
<td>0.16***</td>
<td>0.16***</td>
<td>0.16</td>
</tr>
<tr>
<td>0%-1%</td>
<td>1.20***</td>
<td>0.74***</td>
<td>0.62***</td>
<td>0.48***</td>
<td>0.16***</td>
<td>0.16***</td>
<td>0.76</td>
</tr>
<tr>
<td>1%-2%</td>
<td>0.54***</td>
<td>0.64***</td>
<td>0.46***</td>
<td>0.52***</td>
<td>0.45***</td>
<td>0.26***</td>
<td>0.48</td>
</tr>
<tr>
<td>2%-3%</td>
<td>0.28***</td>
<td>0.11**</td>
<td>0.30***</td>
<td>0.36***</td>
<td>0.40***</td>
<td>0.40***</td>
<td>0.31</td>
</tr>
<tr>
<td>3%-4%</td>
<td>0.02</td>
<td>0.11</td>
<td>0.06</td>
<td>0.05</td>
<td>0.11*</td>
<td>0.17*</td>
<td>0.08</td>
</tr>
<tr>
<td>4%-5%</td>
<td>0.04</td>
<td>0.07</td>
<td>0.01</td>
<td>0.08</td>
<td>0.07*</td>
<td>0.07*</td>
<td>0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ( R^2 )</th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.10</td>
<td>0.11</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>0%-1%</td>
<td>0.52</td>
<td>0.54</td>
<td>0.54</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.51</td>
</tr>
<tr>
<td>1%-2%</td>
<td>0.25</td>
<td>0.49</td>
<td>0.45</td>
<td>0.55</td>
<td>0.55</td>
<td>0.27</td>
<td>0.43</td>
</tr>
<tr>
<td>2%-3%</td>
<td>0.16</td>
<td>0.06</td>
<td>0.28</td>
<td>0.37</td>
<td>0.44</td>
<td>0.45</td>
<td>0.29</td>
</tr>
<tr>
<td>3%-4%</td>
<td>0.00</td>
<td>0.03</td>
<td>0.01</td>
<td>0.07</td>
<td>0.07</td>
<td>0.12</td>
<td>0.04</td>
</tr>
<tr>
<td>4%-5%</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>1 yr</td>
<td>2 yrs</td>
<td>3 yrs</td>
<td>5 yrs</td>
<td>7 yrs</td>
<td>10 yrs</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>Mean</td>
<td>0.97</td>
<td>1.79</td>
<td>2.44</td>
<td>3.36</td>
<td>3.89</td>
<td>4.31</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.80</td>
<td>1.76</td>
<td>2.76</td>
<td>4.58</td>
<td>6.09</td>
<td>8.00</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>1.21</td>
<td>1.01</td>
<td>0.88</td>
<td>0.73</td>
<td>0.64</td>
<td>0.54</td>
</tr>
<tr>
<td><strong>LRSQ(3,1)</strong></td>
<td>Mean</td>
<td>0.40</td>
<td>0.75</td>
<td>1.07</td>
<td>1.63</td>
<td>2.10</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.62</td>
<td>1.31</td>
<td>2.10</td>
<td>3.69</td>
<td>5.05</td>
<td>6.61</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.64</td>
<td>0.57</td>
<td>0.51</td>
<td>0.44</td>
<td>0.42</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>LRSQ(3,2)</strong></td>
<td>Mean</td>
<td>0.39</td>
<td>0.70</td>
<td>0.98</td>
<td>1.49</td>
<td>1.89</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.58</td>
<td>1.23</td>
<td>1.94</td>
<td>3.28</td>
<td>4.46</td>
<td>5.82</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.68</td>
<td>0.57</td>
<td>0.51</td>
<td>0.45</td>
<td>0.42</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>LRSQ(3,3)</strong></td>
<td>Mean</td>
<td>0.27</td>
<td>0.58</td>
<td>0.88</td>
<td>1.39</td>
<td>1.76</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.61</td>
<td>1.21</td>
<td>1.87</td>
<td>3.20</td>
<td>4.37</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.44</td>
<td>0.48</td>
<td>0.47</td>
<td>0.43</td>
<td>0.40</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Table 4: Unconditional excess swap returns.
The table reports annualized means and volatilities of nonoverlapping monthly excess returns on interest rate swaps with a 1-month forward start. Also reported are the annualized Sharpe ratios (SR). Excess returns are in percent. The top panel shows results in the data, where each time series consists of 200 monthly observations from February 1997 to August 2013. The lower panels show results in simulated data, where each time series consists of 600,000 monthly observations (50,000 years).
Table 5: Conditional expected excess swap returns.
The table reports results from regressing nonoverlapping monthly excess swap returns on previous month’s term structure slope and implied volatility (including a constant). Consider the results for the 5-year maturity: The excess return is on an interest rate swap with a 1-month forward start and a 5-year swap maturity. The term structure slope is the difference between the 5-year swap rate and 1-month LIBOR. The implied volatility is the normal implied volatility of a swaption with a 1-month option expiry and a 5-year swap maturity. Excess returns are in percent, and the term structure slopes and implied volatilities are standardized. The top panel shows results in the data, where each time series consists of 200 monthly observations from February 1997 to August 2013. *-statistics, corrected for heteroscedasticity and serial correlation up to 12 lags using the method of Newey and West (1987), are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The lower panels show results in simulated data, where each time series consists of 600,000 monthly observations (50,000 years).
Figure 1: Level-dependence in volatility of 1-year swap rate.
The figure shows the 3-month normal implied volatility of the 1-year swap rate (in basis points) plotted against the level of the 1-year swap rate. The grey area marks the possible range of implied volatilities in case of the $LRSQ(3,3)$ specification.
Figure 2: Data and fit.
Panel A1 shows time series of the 1-year, 5-year, and 10-year swap rates (displayed as thick light-grey, thick dark-grey, and thin black lines, respectively). Panel B1 shows time series of the normal implied volatilities on three “benchmark” swaptions: the 3-month option on the 2-year swap, the 2-year option on the 2-year swap, and the 5-year option on the 5-year swap (displayed as thick light-grey, thick dark-grey, and thin black lines, respectively). Panels A2 and B2 show the fit to swap rates and implied volatilities, respectively, in case of the $LRSQ(3,3)$ specification. Panels A3 and B3 show time series of the root-mean-squared pricing errors (RMSEs) of swap rates and implied volatilities, respectively, in case of the $LRSQ(3,3)$ specification. 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 866 weekly observations from January 29, 1997 to August 28, 2013.
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 866 weekly observations from January 29, 1997 to August 28, 2013.
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 866 weekly observations from January 29, 1997 to August 28, 2013. The grey areas mark the possible range of factor volatilities for a given factor level.
Figure 5: Conditional distribution of short rate.
Conditional on the short rate being between 0 and 25 basis points, Panels A-C display histograms showing the frequency distribution of the future short rate at a 1-year, 2-year, and 5-year horizon, respectively. Panel D displays the mean and median paths of the short rate. The frequency distributions are obtained from 2,600,000 weekly observations (50,000 years) of the short rate simulated from the $LRSQ(3,3)$ specification.
Figure 6: Changing characteristics of “level” factor at the zero lower bound. The figure shows the factor loadings of the first principal component of the term structure when the short rate is away from the zero lower bound (thin black lines) and close to the ZLB (thick grey lines). The factor loadings are constructed from the eigenvector corresponding to the largest eigenvalue of the covariance matrix of weekly changes in swap rates. The solid lines show the factor loadings in the data, where the non-ZLB sample period consists of 620 weekly observations from January 29, 1997 to December 16, 2008, and the ZLB sample period consists of 246 weekly observations from December 16, 2008 to August 28, 2013. The dashed, dashed-dotted, and dotted lines show the factor loadings implied by the \( LRSQ(3,1) \), \( LRSQ(3,2) \), and \( LRSQ(3,3) \) specifications, respectively. The model-implied factor loadings are obtained from simulated data, where each time series consists of 2,600,000 weekly observations (50,000 years). In the simulated data, the non-ZLB (ZLB) sample consists of those observations where the short rate is larger (less) than 25 basis points.
Figure 7: Level-dependence in volatility.
For each swap maturity, weekly changes in the 3-month normal implied volatility of the swap rate are regressed on weekly changes in the level of the swap rate (including a constant). Regressions are run unconditionally as well as conditional on swap rates being in the intervals 0%-1%, 1%-2%, 2%-3%, 3%-4%, and 4%-5%, respectively. Panels A and C show the average (across swap maturities) slope coefficients and $R^2$s, respectively. Panels B and D show the average (across swap maturities) model-implied slope coefficients and $R^2$s, respectively. In each panel, the first bar corresponds to the unconditional regressions, while the second to sixth bars correspond to the conditional regressions. Model-implied values are obtained by running the regressions on data simulated from the $LRSQ(3,3)$ specification, where each time series consists of 2,600,000 weekly observations (50,000 years).
Figure 8: Nonlinearity in short-rate dynamics.
For the $LRSQ(3,0)$ and $A_3(3)$ models, the figure displays the drift and instantaneous variance of $r_t$ as a function of the factors. For both models, 50,000 years of weekly observations of the state variables are simulated. The state vector is set equal to its mean value after which each of the factors are varied between its minimum and maximum value attained in the simulation. Panels A1 and A2 display the drift of $r_t$ under $\mathbb{P}$ (solid lines) and $\mathbb{Q}$ (dash-dotted lines) for the $LRSQ(3,0)$ and $A_3(3)$ models, respectively. Panels B1 and B2 display the instantaneous variance of $r_t$ (multiplied by $10^4$) for the $LRSQ(3,0)$ and $A_3(3)$ models, respectively. Note that in the $A_3(3)$ model, the instantaneous variance of $r_t$ only depends on $X_{3t}$; therefore, the lines corresponding to $X_{1t}$ and $X_{2t}$ lie on top of each other. For expositional ease, the factors are normalized so that they range between zero and one.
References


University.


Online Appendix to
“Linear-Rational Term Structure Models”

Damir Filipović       Martin Larsson       Anders Trolle
EPFL and Swiss Finance Institute
September 4, 2014

Abstract

Section B provides the proofs of all lemmas in the paper. Section C clarifies the connection between bond market completeness and the existence of unspanned factors. Section D gives the exponential-affine transform formula used for swaption pricing. Section E contains the explicit $LRSQ(3,n)$ specifications for $n = 1, 2, 3$. Section F provides summary statistics of the data and details on the estimation approach. Section G supplements Sections 6.5, 6.6, and 6.7 by providing additional details on model performance. Section H supplements Section 6.7 by documenting similar volatility dynamics in a global data set. Section I supplements Section 6.8 by showing that the results are robust to alternative swap return computations. Section J supplements Section 6.9 by giving additional details on parameter estimates and model performance.
B Proofs of Lemmas

This appendix provides the proofs of all lemmas in the paper.

Proof of Lemma A.1

Let \( \{e_1, \ldots, e_d\} \) denote the canonical basis of \( \mathbb{R}^d \). We have for \( p = 0, \ldots, d - 1 \),

\[
\psi^\top \kappa^p S^{-1} e_i = \hat{\psi}^\top \hat{\kappa}^p e_i.
\]

Note that \( \hat{\kappa}^p \) has the same block triangular structure as \( \hat{\kappa} \) for all \( p \geq 1 \). Thus for \( i = m + 1, \ldots, d \), the right side above is zero for all \( p \geq 0 \), so \( S^{-1} e_i \in U \). We deduce that \( S(U) \supseteq \{0\} \times \mathbb{R}^n \) holds, as claimed.

Proof of Lemma A.2

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

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

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

That (i) and (ii) imply (iii) is obvious. Conversely, (iii) implies that

\[
\mathbb{R}^d = \text{span} \{ \psi, \kappa^\top \psi, \ldots, \kappa^{(d-1)\top} \psi \} \subseteq \text{span} \{ \kappa^\top \psi, \ldots, \kappa^{d\top} \psi \},
\]

which yields (i).

Proof of Lemma A.3

First note, plugging (34) in expression (29) for \( \nabla F(\tau, x) \) gives the following identity

\[
\text{span} \{ \nabla F(\tau, x) : \tau \geq 0 \} = \text{span} \left\{ \kappa^p \psi - \frac{\psi^\top \kappa^p (x - \theta)}{\psi^\top x} \psi : p = 1, \ldots, d \right\}. \tag{B.1}
\]

Now assume (i) holds: we can find \( a_1, \ldots, a_d \) such that \( \psi = \sum_{p=1}^{d} a_p \kappa^p \psi \). Together with
the representation (B.1) we deduce that the vector
\[
\sum_{p=1}^{d} a_p \left[ \kappa_p^T \psi - \frac{\psi^T \kappa_p (x - \theta)}{\phi + \psi^T x} \psi \right] = \psi - \frac{\psi^T (x - \theta)}{\phi + \psi^T x} \psi = \frac{\phi + \psi^T \theta}{\phi + \psi^T x} \psi
\]
lies in span \{\nabla F(\tau, x) : \tau \geq 0\}. So does \psi since \phi + \psi^T \theta \neq 0, and it follows from (B.1) again that we have span \{\nabla F(\tau, x) : \tau \geq 0\} \supseteq \text{span}\{\psi, \kappa^T \psi, \ldots, \kappa^d \psi\} = \mathcal{U}^\perp, \text{ see (28)}. Since the other inclusion holds by definition of \mathcal{U}, this proves (ii).

Next assume (ii) holds. To this end, we use that \psi \in \mathcal{U}^\perp together with the representation (B.1) to find \(a_1, \ldots, a_d\) such that
\[
\psi = \sum_{p=1}^{d} a_p \left[ \kappa_p^T \psi - \frac{\psi^T \kappa_p (x - \theta)}{\phi + \psi^T x} \psi \right].
\]
Re-arranging this expression and writing \(u = \sum_{p=1}^{d} a_p \kappa_p^T \psi\) yields
\[
\psi \left( 1 + \frac{u^T (x - \theta)}{\phi + \psi^T x} \right) = u.
\]
It follows that \psi lies in span\{\kappa^T \psi, \ldots, \kappa^d \psi\}, which proves that (i) holds.

**Proof of Lemma A.4**

Consider an arbitrary unspanned direction \(\xi \in \mathcal{U}\). Since \(F(\tau, x + s\xi)\) is constant in \(s\), we have that \(\nabla G(\tau, x)^\top \xi = 0\) if and only if \(\nabla \overline{G}(\tau, x)^\top \xi = 0\), where we define
\[
\overline{G}(\tau, x) = \nabla F(\tau, x)^\top a(x) \nabla F(\tau, x).
\]
Now, for any \(x \in E\) and \(\tau \geq 0\), we have
\[
\frac{d}{ds} \nabla F(\tau, x + s\xi) \bigg|_{s=0} = \nabla \left( \nabla F(\tau, x)^\top \xi \right) = 0.
\]
Hence, by the product rule,
\[
\nabla \overline{G}(\tau, x)^\top \xi = \frac{d}{ds} \overline{G}(\tau, x + s\xi) \bigg|_{s=0} = \frac{d}{ds} \nabla F(\tau, x)^\top a(x + s\xi) \nabla F(\tau, x) \bigg|_{s=0}.
\]
The right side is zero for all \( x \in E, \tau \geq 0 \) if and only if, for every \( x \in E \),

\[
\frac{d}{ds} \eta^\top a(x + s\xi) \bigg|_{s=0} = 0
\]

holds for all \( \eta \in \text{span} \{\nabla F(\tau, x) : \tau \geq 0\} \). But we always have

\[\text{span} \{\nabla F(\tau, x) : \tau \geq 0\} \subseteq \mathcal{U}^\perp,\]

with equality if \( \phi + \psi^\top \theta \neq 0 \) and \( \psi \in \text{span} \{\kappa^\top \psi, \ldots, \kappa^d^\top \psi\} \), due to Lemma A.3. This proves that the inclusion

\[
\mathcal{U} \cap \mathcal{W} \supseteq \left\{ \xi \in \mathcal{U} : \frac{d}{ds} \eta^\top a(x + s\xi) \bigg|_{s=0} = 0 \text{ for all } x \in E, \eta \in \mathcal{U}^\perp \right\} = \mathcal{U} \cap \bigcap_{\eta \in \mathcal{U}^\perp} \ker \eta^\top a(\cdot) \eta
\]

holds, with equality if \( \phi + \psi^\top \theta \neq 0 \) and \( \psi \in \text{span} \{\kappa^\top \psi, \ldots, \kappa^d^\top \psi\} \). Using polarization identity,

\[
4\eta_1^\top a(x)\eta_2 = (\eta_1 + \eta_2)^\top a(x)(\eta_1 + \eta_2) - (\eta_1 - \eta_2)^\top a(x)(\eta_1 - \eta_2)
\]

and the fact that \( \mathcal{U}^\perp \) is a linear space, the lemma follows.

**Proof of Lemma A.5**

We first prove the result for \( \rho = 0 \). An application of Itô’s formula shows that the process

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

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

\[
Y_t = Y_0 + e^{-\kappa(T-t)}M_t - \int_0^t M_s e^{-\kappa(T-s)}ds,
\]
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 $0 \leq t \leq u$,

$$
\mathbb{E}_t[Y_u] = Y_0 + e^{-\kappa(T-u)}M_t - \int_0^u M_{s\wedge t} \kappa e^{-\kappa(T-s)}ds
$$

$$
= Y_t + M_t \left[ e^{-\kappa(T-u)} - e^{-\kappa(T-t)} - \int_t^u \kappa e^{-\kappa(T-s)}ds \right] = Y_t,
$$

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

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

as claimed. If $\rho$ is a bounded stopping time, then the $L^1$-boundedness of $Y$, and hence of $X$, implies that $X_\rho$ 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 Lemma A.6**

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

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

so that, by Lemma A.5, $\mathbb{E}_\rho[X_{\rho+\tau}] = G(\tau, X_\rho)$ holds for any bounded stopping time $\rho$ 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 $G_i(\tau, \lambda e_j) = 0 + \int_0^\tau \partial_\tau G_i(s, \lambda e_j)ds \leq -2\tau$ for all $\tau \in [0, 2\varepsilon]$. By continuity of $(\tau, x) \mapsto G(\tau, x)$ there is some $r > 0$ such that

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

5
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

$$
\mathbb{E}[1_A X_{i, \rho + \varepsilon}] = \mathbb{E}[1_A \mathbb{E}_\rho [X_{i, \rho + \varepsilon}]] = \mathbb{E}[1_A G_i(\varepsilon, X_\rho)] \leq \varepsilon \mathbb{P}(A) < 0,
$$

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

**Proof of Lemma A.7**

Lemma A.6 implies that $-1_p^T \kappa_i \geq 0$ for $i > p$. Since the state components $x_i$ are nonnegative and unbounded from above, it is then obvious that $\alpha^*$ is finite if and only if $1_p^T \kappa_i = 0$ for $i > p$. This means that the submatrix $\kappa_{1..p, p+1..d}$ is zero.

The expressions for $\alpha^*$ and $\alpha_*$ now follow by observing that for each $x \in \mathbb{R}^d_+$, $\rho(x)$ is a convex combination of the elements in $S_p$. 
C Bond Market Completeness

We discuss how the notion of spanning relates to bond market completeness in the linear-rational diffusion model of Section 3. The following definition of completeness is standard.

**Definition C.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 C.2.** Assume that the filtration $\mathcal{F}_t$ is generated by the Brownian motion $B_t$, that $\sigma(X_t)$ is invertible $dt \otimes d\mathbb{P}$-almost surely, and that $\phi + \psi^\top \theta \neq 0$. Then the following conditions are equivalent:

(i) bond markets are complete;

(ii) the term structure $F(\tau,x)$ is injective;

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

(iv) $\text{span}\{\nabla F(\tau,X_t) : \tau \geq 0\} = \mathbb{R}^d$, $dt \otimes d\mathbb{P}$-almost surely.

The proof of Theorem C.2 requires some notation and the following additional lemma. For a multiindex $k = (k_1, \ldots, k_d) \in \mathbb{N}_0^d$ we write $|k| = k_1 + \cdots + k_d$, $x^k = x_1^{k_1} \cdots x_d^{k_d}$, and $\partial^k = \partial^{|k|}/\partial x_1^{k_1} \cdots \partial x_d^{k_d}$.

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

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

up to a $dt \otimes d\mathbb{P}$-nullset.
Proof. Let \( n \geq 0 \) and suppose we have, up to a nullset,
\[
\{ f(t, X_t) = 0 \} \subseteq \{ \partial^k f(t, X_t) = 0 \}
\] (C.1)
for all \( k \in \mathbb{N}_0^d \) with \( |k| = n \). Fix such a \( k \) and set \( g(t, x) = \partial^k f(t, x) \). The occupation time formula, see Revuz and Yor (1999, Corollary VI.1.6), yields
\[
1_{\{g(t,X_t) = 0\}} \nabla g(t,X_t)^\top a(X_t) \nabla g(t,X_t) = 0 \ dt \otimes dP \text{-a.s.}
\]
This implies \( \{g(t, X_t) = 0\} \subseteq \{ \nabla g(t, X_t)^\top a(X_t) \nabla g(t, X_t) = 0\} \) up to a nullset. Since \( a(X_t) \) is invertible \( dt \otimes dP \)-almost surely we get, again up to a nullset,
\[
\{g(t, X_t) = 0\} \subseteq \{ \nabla g(t, X_t) = 0\}.
\]
We deduce that (C.1) holds for all \( k \in \mathbb{N}_0^d \) with \( |k| = n + 1 \). Since (C.1) is trivially true for \( n = 0 \), the result follows by induction. \( \square \)

Proof of Theorem C.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 \( T_i \geq T, i = 1, \ldots, m \), such that \( dt \otimes dP \)-almost surely the volatility vectors \( \nu(T_i - t, X_t) \) span \( \mathbb{R}^d \). Since \( \sigma(X_t) \) is invertible \( dt \otimes dP \)-almost surely this happens if and only if \( dt \otimes dP \)-almost surely the vectors \( \nabla F(T_i - t, X_t) \) span \( \mathbb{R}^d \). This shows, in particular, that (i) implies (iv).

To prove that (iv) implies (iii), suppose (iii) fails. Lemma A.3 then implies that for each \( x \in E \), span\{\nabla F(\tau, x) : \tau \geq 0\} is not all of \( U^\perp = \mathbb{R}^d \). Thus (iv) fails.

Equivalence of (iii) and (ii) follows easily from Theorem 2.5.

It remains to prove that (iii) implies (i), so we assume that \( \kappa \) is invertible and \( U = \{0\} \). Choose maturities \( T_1, \ldots, T_d \) greater than or equal to \( T \) so that the function
\[
g(t, x) = \det \left( \nabla F(T_1 - t, x) \cdots \nabla F(T_d - t, x) \right)
\]
is not identically zero. This is possible by Lemma A.3 (ii). A calculation yields
\[
\nabla F(\tau, x) = \frac{e^{-\alpha \tau}}{(\psi + \psi^\top x)^2} \eta(\tau, x),
\]
where \( \eta(\tau, x) \) is a vector of first degree polynomials in \( x \) whose coefficients are analytic.
functions of $\tau$. Defining
\[ f(t, x) = \det \left( \eta(T_1 - t, x) \cdots \eta(T_d - t, x) \right), \]
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 $dt \otimes d\mathbb{P}$-nullset. Indeed, then $\{ g(t, X_t) = 0 \}$ is also a $dt \otimes d\mathbb{P}$-nullset, implying that completeness holds. To prove that $\{ f(t, X_t) = 0 \}$ is a $dt \otimes d\mathbb{P}$-nullset, note that $f(t, x)$ is of the form
\[ f(t, x) = \sum_{|k| \leq n} c_k(t) x^k, \]
where $n = \max_{0 \leq t \leq T} \deg f(t, \cdot) < \infty$. Lemma C.3 implies
\[ \{ f(t, X_t) = 0 \} \subseteq \bigcap_{|k|=n} \{ c_k(t) = 0 \}, \]
up to a nullset. Assume for contradiction that the left side is of positive $dt \otimes d\mathbb{P}$-measure. Then so is the right side, whence all the $c_k$ (which are deterministic) vanish on a $t$-set of positive Lebesgue measure. The zero set of each $c_k$ 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 $\max_{0 \leq t \leq T} \deg f(t, \cdot) \leq n - 1$ (if $n \geq 1$) 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 $dt \otimes d\mathbb{P}$-nullset, as required. Theorem C.2 is thus proved. \[ \square \]
D Exponential-Affine Transform Formula

Let $X_t$ be the square-root process (23). We reproduce here the exponential-affine transform formula that is available for $X_t$. It can be found in, e.g., Duffie, Pan, and Singleton (2000) and Filipović (2009, Theorem 10.3).

Lemma D.1 (Exponential-Affine Transform Formula). For any $0 \leq t \leq T$, and $u \in \mathbb{C}$, $v \in \mathbb{C}^d$ such that $\mathbb{E}[|\exp(v^\top X_T)|] < \infty$ we have

$$\mathbb{E}_t \left[ e^{u + v^\top X_T} \right] = e^{\Phi(T-t) + \Psi(T-t)^\top X_t},$$

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

$$\Phi'(\tau) = b^\top \Psi(\tau)$$

$$\Psi_i'(\tau) = \beta_i^\top \Psi(\tau) + \frac{1}{2} \sigma_i^2 \Psi_i(\tau)^2, \quad i = 1, \ldots, d,$$

with initial condition $\Phi(0) = u$, $\Psi(0) = v$. The solution to this system is unique.
E Explicit $LRSQ(3,n)$ Specifications

This section contains the explicit $LRSQ(3,n)$ specifications for $n = 1, 2, 3$.

$LRSQ(3,1)$

The mean reversion matrix is of the form

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

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

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

The corresponding volatility matrix is

$$\hat{\sigma}(z, u) = \begin{pmatrix} \sigma_1 \sqrt{z_1 - u_1} & 0 & 0 & \sigma_4 \sqrt{u_1} \\ 0 & \sigma_2 \sqrt{z_2} & 0 & 0 \\ 0 & 0 & \sigma_3 \sqrt{z_3} & 0 \\ 0 & 0 & 0 & \sigma_4 \sqrt{u_1} \end{pmatrix}.$$
**LRSQ(3,2)**

The mean reversion matrix is of the form

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

The term structure factors and unspanned factors become

\[
\begin{pmatrix}
Z_{1t} \\
Z_{2t} \\
Z_{3t} \\
U_{1t} \\
U_{2t}
\end{pmatrix} = SX_t = \begin{pmatrix}
X_{1t} + X_{4t} \\
X_{2t} + X_{5t} \\
X_{3t} \\
X_{4t} \\
X_{5t}
\end{pmatrix},
\]

and the transformed mean reversion matrix is given by

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

The corresponding volatility matrix is

\[
\hat{\sigma}(z, u) = \begin{pmatrix}
\sigma_1 \sqrt{z_1 - u_1} & 0 & 0 & \sigma_4 \sqrt{u_1} & 0 \\
0 & \sigma_2 \sqrt{z_2 - u_2} & 0 & 0 & \sigma_5 \sqrt{u_2} \\
0 & 0 & \sigma_3 \sqrt{z_3} & 0 & 0 \\
0 & 0 & 0 & \sigma_4 \sqrt{u_1} & 0 \\
0 & 0 & 0 & 0 & \sigma_5 \sqrt{u_2}
\end{pmatrix}.
\]
The mean reversion matrix is of the form
\[
\kappa = \begin{pmatrix}
\kappa_{11} & \kappa_{12} & \kappa_{13} & 0 & 0 & 0 \\
\kappa_{21} & \kappa_{22} & \kappa_{23} & 0 & 0 & 0 \\
\kappa_{31} & \kappa_{32} & \kappa_{33} & 0 & 0 & 0 \\
0 & 0 & 0 & \kappa_{11} & \kappa_{12} & \kappa_{13} \\
0 & 0 & 0 & \kappa_{21} & \kappa_{22} & \kappa_{23} \\
0 & 0 & 0 & \kappa_{31} & \kappa_{32} & \kappa_{33}
\end{pmatrix}.
\]

The term structure factors and unspanned factors become
\[
\begin{pmatrix}
Z_{1t} \\
Z_{2t} \\
Z_{3t} \\
U_{1t} \\
U_{2t} \\
U_{3t}
\end{pmatrix} = S_X t = \begin{pmatrix}
X_{1t} + X_{4t} \\
X_{2t} + X_{5t} \\
X_{3t} + X_{6t} \\
X_{4t} \\
X_{5t} \\
X_{6t}
\end{pmatrix},
\]
and the transformed mean reversion matrix is given by
\[
\hat{\kappa} = \begin{pmatrix}
\kappa_{11} & \kappa_{12} & \kappa_{13} & 0 & 0 & 0 \\
\kappa_{21} & \kappa_{22} & \kappa_{23} & 0 & 0 & 0 \\
\kappa_{31} & \kappa_{32} & \kappa_{33} & 0 & 0 & 0 \\
0 & 0 & 0 & \kappa_{11} & \kappa_{12} & \kappa_{13} \\
0 & 0 & 0 & \kappa_{21} & \kappa_{22} & \kappa_{23} \\
0 & 0 & 0 & \kappa_{31} & \kappa_{32} & \kappa_{33}
\end{pmatrix}.
\]
The corresponding volatility matrix is

$$\hat{\sigma}(z,u) = \begin{pmatrix}
\sigma_1 \sqrt{z_1 - u_1} & 0 & 0 & \sigma_1 \sqrt{u_1} & 0 & 0 \\
0 & \sigma_2 \sqrt{z_2 - u_2} & 0 & 0 & \sigma_5 \sqrt{u_2} & 0 \\
0 & 0 & \sigma_3 \sqrt{z_3 - u_3} & 0 & 0 & \sigma_6 \sqrt{u_3} \\
0 & 0 & 0 & \sigma_4 \sqrt{u_1} & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_5 \sqrt{u_2} & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_6 \sqrt{u_3}
\end{pmatrix}.$$
Summary Statistics and Estimation Approach

Summary Statistics of Data
Table F.1 shows summary statistics of swap rates and NIVs. The term structure of swap rates is upward-sloping, on average, while the standard deviations of swap rates decrease with maturity. Along the swap maturity dimension, the term structure of NIVs is hump-shaped, on average, for short and intermediate option expiries and downward-sloping for the longest option expiry. Along the option expiry dimension, the term structure of NIVs is upward-sloping, on average, for the shortest swap maturity, hump-shaped for intermediate swap maturities, and downward-sloping for the longest swap maturity. The standard deviations of NIVs tend to decrease with both swap maturity and option expiry.

Quasi-Maximum Likelihood Estimation
We estimate by quasi-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 observable swap rates and NIVs, 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 NIVs observed at time $t$. The measurement equation is given by

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

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

While the transition density of $X_t$ is unknown, its conditional mean and variance is known in closed form, because $X_t$ is a square-root 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

$$X_t = \Phi_0 + \Phi_X X_{t-1} + w_t, \quad w_t \sim N(0, Q_t),$$
where $Q_t$ is a linear function of $X_{t-1}$.

As both swap rates and NIVs are nonlinearly related to the state variables, we apply the nonlinear unscented Kalman filter.\(^1\) 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

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

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

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

Approximating the true transition density with a Gaussian makes this a quasi-maximum likelihood (QML) procedure. While QML estimation has been shown to be consistent in many settings, it is in fact not consistent in a Kalman filter setting, because the conditional covariance matrix $Q_t$ in the recursions depends on the Kalman filter estimates of the state variables instead of the true, but unobservable, values; see, e.g., Duan and Simonato (1999). However, simulation results in several papers have shown this issue to be negligible in practice.

\(^1\)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.
<table>
<thead>
<tr>
<th></th>
<th>Swap maturity</th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swaps</td>
<td></td>
<td>3.19</td>
<td>3.46</td>
<td>3.72</td>
<td>4.13</td>
<td>4.43</td>
<td>4.72</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>78.5</td>
<td>92.6</td>
<td>97.6</td>
<td>105.0</td>
<td>105.9</td>
<td>105.6</td>
</tr>
<tr>
<td>1 yr</td>
<td></td>
<td>97.1</td>
<td>101.2</td>
<td>103.2</td>
<td>106.0</td>
<td>106.1</td>
<td>104.6</td>
</tr>
<tr>
<td>2 yrs</td>
<td></td>
<td>106.2</td>
<td>106.5</td>
<td>106.9</td>
<td>107.0</td>
<td>106.0</td>
<td>103.4</td>
</tr>
<tr>
<td>5 yrs</td>
<td></td>
<td>107.9</td>
<td>106.2</td>
<td>105.0</td>
<td>102.4</td>
<td>99.1</td>
<td>97.1</td>
</tr>
<tr>
<td><strong>Panel B: Standard deviation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swaps</td>
<td></td>
<td>2.24</td>
<td>2.14</td>
<td>2.03</td>
<td>1.83</td>
<td>1.67</td>
<td>1.52</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>30.9</td>
<td>34.6</td>
<td>33.7</td>
<td>31.7</td>
<td>28.8</td>
<td>27.1</td>
</tr>
<tr>
<td>1 yr</td>
<td></td>
<td>30.8</td>
<td>30.2</td>
<td>27.8</td>
<td>24.5</td>
<td>22.0</td>
<td>19.9</td>
</tr>
<tr>
<td>2 yrs</td>
<td></td>
<td>26.1</td>
<td>24.4</td>
<td>22.5</td>
<td>19.8</td>
<td>18.3</td>
<td>16.7</td>
</tr>
<tr>
<td>5 yrs</td>
<td></td>
<td>14.8</td>
<td>14.0</td>
<td>13.4</td>
<td>12.4</td>
<td>11.5</td>
<td>11.2</td>
</tr>
<tr>
<td><strong>Panel C: Minimum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swaps</td>
<td></td>
<td>0.30</td>
<td>0.34</td>
<td>0.43</td>
<td>0.74</td>
<td>1.14</td>
<td>1.55</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>17.3</td>
<td>18.0</td>
<td>24.3</td>
<td>40.2</td>
<td>51.9</td>
<td>57.2</td>
</tr>
<tr>
<td>1 yr</td>
<td></td>
<td>24.2</td>
<td>25.9</td>
<td>35.4</td>
<td>52.0</td>
<td>63.9</td>
<td>63.3</td>
</tr>
<tr>
<td>2 yrs</td>
<td></td>
<td>39.1</td>
<td>43.6</td>
<td>52.0</td>
<td>64.8</td>
<td>70.0</td>
<td>68.4</td>
</tr>
<tr>
<td>5 yrs</td>
<td></td>
<td>73.1</td>
<td>72.0</td>
<td>72.7</td>
<td>71.3</td>
<td>71.1</td>
<td>69.7</td>
</tr>
<tr>
<td><strong>Panel D: Maximum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swaps</td>
<td></td>
<td>7.51</td>
<td>7.65</td>
<td>7.70</td>
<td>7.75</td>
<td>7.77</td>
<td>7.80</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>224.4</td>
<td>212.5</td>
<td>208.2</td>
<td>203.4</td>
<td>206.2</td>
<td>208.3</td>
</tr>
<tr>
<td>1 yr</td>
<td></td>
<td>169.5</td>
<td>169.1</td>
<td>173.3</td>
<td>175.2</td>
<td>174.0</td>
<td>170.8</td>
</tr>
<tr>
<td>2 yrs</td>
<td></td>
<td>159.8</td>
<td>160.7</td>
<td>160.4</td>
<td>159.9</td>
<td>157.7</td>
<td>154.4</td>
</tr>
<tr>
<td>5 yrs</td>
<td></td>
<td>153.9</td>
<td>142.0</td>
<td>136.5</td>
<td>133.5</td>
<td>130.5</td>
<td>127.4</td>
</tr>
</tbody>
</table>

Table F.1: Summary statistics.
The table reports the mean, standard deviation, minimum, and maximum of each time series. Statistics for swap rates are reported in percentages. Statistics for swaption normal implied volatilities are reported in basis points. Each time series consists of 866 weekly observations from January 29, 1997 to August 28, 2013.
Details on Model Performance

The aggregate RMSEs reported in Table 2 in the paper have the potential to mask problems with fitting individual swap rates or swaption implied volatilities. Therefore, Table G.1 shows the RMSEs for individual swap rates and NIVs. Generally, the fit is fairly uniform across the swap term structure and across the implied volatility surface, with the possible exception of the 3-month option on the 1-year swap rate. Taking the $LRSQ(3, 3)$ specification as an example, the RMSEs for swap rates vary between 3.52 bps (for the 7-year swap rate) and 4.69 bps (for the 5-year swap rate). The RMSEs for NIVs vary between 2.84 bps (for the 2-year option on the 5-year swap rate) and 11.18 bps (for the 3-month option on the 1-year swap rate). It is common to observe that the fit to implied volatilities deteriorates at the edges of an implied volatility surface.

Figure G.1 (for the $LRSQ(3, 1)$ specification) and Figure G.2 (for the $LRSQ(3, 2)$ specification) are the counterparts to Figure 5 in the paper. The conditional distributions generated by the $LRSQ(3, 2)$ specification are similar to those generated by the $LRSQ(3, 3)$ specification, while the $LRSQ(3, 1)$ specification generates somewhat more persistently low short rates.

Figure G.3 (for the $LRSQ(3, 1)$ specification) and Figure G.4 (for the $LRSQ(3, 2)$ specification) are the counterparts to Figure 7 in the paper. The $LRSQ(3, 1)$ specification generates too high a degree of level-dependence in volatility, both unconditionally and conditionally, because it only includes one USV factor. The performance of the $LRSQ(3, 2)$ specification is broadly similar to that of the $LRSQ(3, 3)$ specification, generating slightly more level-dependence in volatility at low swap rates and slightly less level-dependence at higher swap rates.
<table>
<thead>
<tr>
<th>Panel A: LRSQ(3,1)</th>
<th>Swap maturity</th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swaps</td>
<td></td>
<td>10.74</td>
<td>7.77</td>
<td>7.21</td>
<td>7.06</td>
<td>6.68</td>
<td>8.21</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>13.62</td>
<td>8.45</td>
<td>7.59</td>
<td>8.74</td>
<td>8.77</td>
<td>9.65</td>
</tr>
<tr>
<td>Swaptions 1 yr</td>
<td></td>
<td>9.35</td>
<td>6.29</td>
<td>4.64</td>
<td>4.74</td>
<td>4.06</td>
<td>5.46</td>
</tr>
<tr>
<td>Swaptions 2 yrs</td>
<td></td>
<td>7.82</td>
<td>5.24</td>
<td>4.90</td>
<td>4.33</td>
<td>4.09</td>
<td>6.48</td>
</tr>
<tr>
<td>Swaptions 5 yrs</td>
<td></td>
<td>6.99</td>
<td>6.48</td>
<td>5.94</td>
<td>4.84</td>
<td>5.67</td>
<td>8.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: LRSQ(3,2)</th>
<th>Swap maturity</th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swaps</td>
<td></td>
<td>4.52</td>
<td>4.07</td>
<td>5.01</td>
<td>4.76</td>
<td>3.36</td>
<td>5.70</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>14.86</td>
<td>7.14</td>
<td>6.02</td>
<td>6.10</td>
<td>6.97</td>
<td>11.22</td>
</tr>
<tr>
<td>Swaptions 1 yr</td>
<td></td>
<td>7.16</td>
<td>6.35</td>
<td>5.71</td>
<td>4.55</td>
<td>4.08</td>
<td>6.81</td>
</tr>
<tr>
<td>Swaptions 2 yrs</td>
<td></td>
<td>6.39</td>
<td>4.63</td>
<td>3.84</td>
<td>2.69</td>
<td>3.06</td>
<td>5.35</td>
</tr>
<tr>
<td>Swaptions 5 yrs</td>
<td></td>
<td>5.06</td>
<td>4.15</td>
<td>3.97</td>
<td>4.22</td>
<td>4.70</td>
<td>6.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: LRSQ(3,3)</th>
<th>Swap maturity</th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swaps</td>
<td></td>
<td>4.62</td>
<td>4.68</td>
<td>4.07</td>
<td>4.69</td>
<td>3.52</td>
<td>4.46</td>
</tr>
<tr>
<td>Swaptions 3 mths</td>
<td></td>
<td>11.18</td>
<td>7.12</td>
<td>6.24</td>
<td>6.29</td>
<td>6.81</td>
<td>9.51</td>
</tr>
<tr>
<td>Swaptions 1 yr</td>
<td></td>
<td>6.70</td>
<td>5.90</td>
<td>4.72</td>
<td>4.06</td>
<td>2.93</td>
<td>4.63</td>
</tr>
<tr>
<td>Swaptions 2 yrs</td>
<td></td>
<td>6.29</td>
<td>4.62</td>
<td>3.66</td>
<td>2.84</td>
<td>2.85</td>
<td>4.52</td>
</tr>
<tr>
<td>Swaptions 5 yrs</td>
<td></td>
<td>4.61</td>
<td>4.09</td>
<td>3.89</td>
<td>3.69</td>
<td>3.92</td>
<td>4.74</td>
</tr>
</tbody>
</table>

Table G.1: Individual RMSEs.
The table reports root mean squared pricing errors (RMSEs) of individual swap rates and normal implied swaption volatilities. Units are basis points. The sample period consists of 866 weekly observations from January 29, 1997 to August 28, 2013.
Figure G.1: Conditional distribution of short rate, $LRSQ(3,1)$
Conditional on the short rate being between 0 and 25 basis points, Panels A-C display histograms showing the frequency distribution of the future short rate at a 1-year, 2-year, and 5-year horizon, respectively. Panel D displays the mean and median paths of the short rate. The frequency distributions are obtained from 2,600,000 weekly observations (50,000 years) of the short rate simulated from the $LRSQ(3,1)$ specification.
Figure G.2: Conditional distribution of short rate, $LRSQ(3,2)$
Conditional on the short rate being between 0 and 25 basis points, Panels A-C display histograms showing the frequency distribution of the future short rate at a 1-year, 2-year, and 5-year horizon, respectively. Panel D displays the mean and median paths of the short rate. The frequency distributions are obtained from 2,600,000 weekly observations (50,000 years) of the short rate simulated from the $LRSQ(3,2)$ specification.
Figure G.3: Level-dependence in volatility, $LRSQ(3,1)$. For each swap maturity, weekly changes in the 3-month normal implied volatility of the swap rate are regressed on weekly changes in the level of the swap rate (including a constant). Regressions are run unconditionally as well as conditional on swap rates being in the intervals 0%-1%, 1%-2%, 2%-3%, 3%-4%, and 4%-5%, respectively. Panels A and C show the average (across swap maturities) slope coefficients and $R^2$s, respectively. Panels B and D show the average (across swap maturities) model-implied slope coefficients and $R^2$s, respectively. In each panel, the first bar corresponds to the unconditional regressions, while the second to sixth bars correspond to the conditional regressions. Model-implied values are obtained by running the regressions on data simulated from the $LRSQ(3,1)$ specification, where each time series consists of 2,600,000 weekly observations (50,000 years).
Figure G.4: Level-dependence in volatility, \( LRSQ(3,2) \).
For each swap maturity, weekly changes in the 3-month normal implied volatility of the swap rate are regressed on weekly changes in the level of the swap rate (including a constant). Regressions are run unconditionally as well as conditional on swap rates being in the intervals 0%-1%, 1%-2%, 2%-3%, 3%-4%, and 4%-5%, respectively. Panels A and C show the average (across swap maturities) slope coefficients and \( R^2 \)'s, respectively. Panels B and D show the average (across swap maturities) model-implied slope coefficients and \( R^2 \)'s, respectively. In each panel, the first bar corresponds to the unconditional regressions, while the second to sixth bars correspond to the conditional regressions. Model-implied values are obtained by running the regressions on data simulated from the \( LRSQ(3,2) \) specification, where each time series consists of 2,600,000 weekly observations (50,000 years).
H Analysis of Volatility Dynamics

In Figure 1 in the paper, we focus on the 1-year swap rate, since this is the swap rate that is closest to the ZLB during the sample period. Here we show that a similar pattern holds true for longer swap maturities. We merge data from the four largest swap markets—the U.S., the Eurozone, the U.K, and Japan—in order to increase the number of data points with very low interest rates (Japanese data is obviously particularly important for this purpose).\(^2\) For the 1-year, 3-year, 5-year, and 10-year swap maturities, Figure H.1 shows a scatterplot of the 3-month NIV of the swap rate (in bps) against the level of the swap rate (data for the 2-year and 7-year swap maturities look very similar). In all cases, we observe that volatility becomes compressed and gradually more level-dependent as the underlying swap rate approaches the ZLB.

\(^2\)As of August 30, 2013, the Depository Trust and Clearing Corporation (DTCC)’s Global Trade Repository contains 220,116 swaption contracts with a combined gross notional value of 29.556 trillion USD equivalent. Of the gross notional value, 42.7%, 33.5%, 13.7%, and 7.2% is denominated in EUR, USD, JPY, and GBP, respectively.
Figure H.1: Level-dependence in volatility
For each swap maturity, the figure shows the 3-month normal implied volatility of the swap rate (in basis points) plotted against the level of the swap rate. Panels A, B, C, and D correspond to swap maturities of 1 year, 3 years, 5 years, and 10 years, respectively. Data from the U.S., the Eurozone, the U.K, and Japan are marked on a scale from light-grey to dark-grey.
I Alternative Swap Return Computations

In the paper, we work with excess returns on forward-starting swaps. Here we show that very similar results are obtained with excess returns on spot-starting swaps as well as zero-coupon bonds bootstrapped from swap rates. When computing returns on spot-starting swaps, we again consider the swap contract described in Section 2.3 that lasts from $T_0$ to $T_n$. The strategy enters into the spot-starting swap contract at time $T_0$, receiving fixed and paying floating. At time $T_0 + \Delta < T_1$, the value of the swap is

$$\Pi_{T_0+\Delta}^{swap} = -P(T_0 + \Delta, T_1)\Delta L(T_0, T_1) - \left( P(T_0 + \Delta, T_1) - P(T_0 + \Delta, T_n) \right)$$

$$+ S_{T_0,T_0}^\Delta \left( \sum_{i=1}^n \Delta P(T_0 + \Delta, T_i) \right).$$

The first term is the value of the first floating-rate payment occurring time $T_1$, the second term is the value of the remaining floating-rate payments, and the last term is the value of the fixed-rate payments.\(^3\) Since the swap has zero initial value, the excess return over a holding period of $\Delta$ is

$$R_{T_0,T_0+\Delta}^e = \frac{\Pi_{T_0+\Delta}^{swap}}{C_{T_0}},$$

where $C_{T_0}$ is the initial amount of capital which earns the risk-free rate. As in the paper, we consider “fully collateralized” swap positions.

Computing returns on zero-coupon bonds bootstrapped from swap rates is straightforward. Buying a zero-coupon bond at $T_0$ with maturity at $T_n$ and holding it over a period of $\Delta$ yields an excess return of

$$R_{T_0,T_0+\Delta}^e = \frac{P(T_0 + \Delta, T_n) - P(T_0, T_n)}{P(T_0, T_n)} - \Delta L(T_0, T_0 + \Delta),$$

Tables I.1 and I.2 are similar to Table 4 in the paper, but based on returns computed from spot-starting swaps and zero-coupon bonds, respectively. In the data, Sharpe ratios inferred from spot-starting swaps are lower than those inferred from forward-starting swaps, particularly for short swap maturities. Sharpe ratios inferred from zero-coupon bonds lie between those inferred from spot-starting swaps and forward-starting swaps. The models

\(^3\)The valuation equation assumes that the payments on the fixed and floating legs occur at the same frequency. When computing returns, we take into account that the fixed-leg payments occur at a semi-annual frequency, while floating-leg payments occur at a quarterly frequency.
tend to capture the risk and return characteristics of spot-starting swaps and zero-coupon bonds even better than is the case for forward-starting swaps.

Tables I.3 and I.4 are similar to Table 5 in the paper, but based on returns computed from spot-starting swaps and zero-coupon bonds, respectively. Generally, the results are very similar across forward-starting swaps, spot-starting swaps, and zero-coupon bonds.
<table>
<thead>
<tr>
<th></th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.39</td>
<td>1.32</td>
<td>2.08</td>
<td>3.10</td>
<td>3.70</td>
<td>4.23</td>
</tr>
<tr>
<td>Vol</td>
<td>0.62</td>
<td>1.60</td>
<td>2.62</td>
<td>4.46</td>
<td>5.99</td>
<td>7.92</td>
</tr>
<tr>
<td>SR</td>
<td>0.63</td>
<td>0.83</td>
<td>0.79</td>
<td>0.69</td>
<td>0.62</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>LRSQ(3,1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.30</td>
<td>0.65</td>
<td>0.97</td>
<td>1.54</td>
<td>2.02</td>
<td>2.57</td>
</tr>
<tr>
<td>Vol</td>
<td>0.47</td>
<td>1.17</td>
<td>1.98</td>
<td>3.60</td>
<td>4.99</td>
<td>6.57</td>
</tr>
<tr>
<td>SR</td>
<td>0.63</td>
<td>0.55</td>
<td>0.49</td>
<td>0.43</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td><strong>LRSQ(3,2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.29</td>
<td>0.60</td>
<td>0.88</td>
<td>1.39</td>
<td>1.80</td>
<td>2.25</td>
</tr>
<tr>
<td>Vol</td>
<td>0.44</td>
<td>1.13</td>
<td>1.86</td>
<td>3.24</td>
<td>4.43</td>
<td>5.81</td>
</tr>
<tr>
<td>SR</td>
<td>0.64</td>
<td>0.53</td>
<td>0.47</td>
<td>0.43</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td><strong>LRSQ(3,3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.21</td>
<td>0.52</td>
<td>0.82</td>
<td>1.33</td>
<td>1.71</td>
<td>2.07</td>
</tr>
<tr>
<td>Vol</td>
<td>0.46</td>
<td>1.08</td>
<td>1.77</td>
<td>3.13</td>
<td>4.33</td>
<td>5.74</td>
</tr>
<tr>
<td>SR</td>
<td>0.45</td>
<td>0.48</td>
<td>0.46</td>
<td>0.43</td>
<td>0.39</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table I.1: Unconditional excess swap returns, spot-starting swaps.
The table reports the annualized means and volatilities of nonoverlapping monthly excess returns on spot-starting interest rate swaps. Also reported are the annualized Sharpe ratios (SR). Excess returns are in percent. The top panel shows results in the data, where each time series consists of 200 monthly observations from February 1997 to August 2013. The lower panels show results in simulated data, where each time series consists of 600,000 monthly observations (50,000 years).
### Table I.2: Unconditional excess returns, zero-coupon bonds.

The table reports the annualized means and volatilities of nonoverlapping monthly excess returns on zero-coupon bonds bootstrapped from swap rates. Also reported are the annualized Sharpe ratios (SR). Excess returns are in percent. The top panel shows results in the data, where each time series consists of 200 monthly observations from February 1997 to August 2013. The lower panels show results in simulated data, where each time series consists of 600,000 monthly observations (50,000 years).

<table>
<thead>
<tr>
<th></th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.58</td>
<td>1.56</td>
<td>2.39</td>
<td>3.61</td>
<td>4.46</td>
<td>5.43</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.71</td>
<td>1.72</td>
<td>2.82</td>
<td>4.96</td>
<td>6.96</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.82</td>
<td>0.91</td>
<td>0.85</td>
<td>0.73</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>LRSQ(3,1)</strong></td>
<td>Mean</td>
<td>0.37</td>
<td>0.74</td>
<td>1.10</td>
<td>1.77</td>
<td>2.39</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.57</td>
<td>1.28</td>
<td>2.14</td>
<td>4.02</td>
<td>5.83</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.64</td>
<td>0.58</td>
<td>0.51</td>
<td>0.44</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>LRSQ(3,2)</strong></td>
<td>Mean</td>
<td>0.37</td>
<td>0.70</td>
<td>1.01</td>
<td>1.60</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.53</td>
<td>1.21</td>
<td>1.97</td>
<td>3.54</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.69</td>
<td>0.58</td>
<td>0.51</td>
<td>0.45</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>LRSQ(3,3)</strong></td>
<td>Mean</td>
<td>0.25</td>
<td>0.58</td>
<td>0.91</td>
<td>1.53</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>Vol</td>
<td>0.57</td>
<td>1.19</td>
<td>1.92</td>
<td>3.51</td>
<td>5.06</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.43</td>
<td>0.48</td>
<td>0.47</td>
<td>0.44</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Table I.3: Conditional excess swap returns, spot-starting swaps.

The table reports results from regressing nonoverlapping monthly excess swap returns on previous month’s term structure slope and implied volatility (including a constant). Consider the results for the 5-year maturity: The excess return is on a spot-starting interest rate swap with a 5-year maturity. The term structure slope is the difference between the 5-year swap rate and 1-month LIBOR. The implied volatility is the normal implied volatility of a swaption with a 1-month option expiry and a 5-year swap maturity. Excess returns are in percent, and the term structure slopes and implied volatilities are standardized. The top panel shows results in the data, where each time series consists of 200 monthly observations from February 1997 to August 2013. \( t \)-statistics, corrected for heteroscedasticity and serial correlation up to 12 lags using the method of Newey and West (1987), are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The lower panels show results in simulated data, where each time series consists of 600,000 monthly observations (50,000 years).

<table>
<thead>
<tr>
<th></th>
<th>1 yr</th>
<th>2 yrs</th>
<th>3 yrs</th>
<th>5 yrs</th>
<th>7 yrs</th>
<th>10 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Slp}} )</td>
<td>-0.017</td>
<td>0.003</td>
<td>0.038</td>
<td>0.097</td>
<td>0.117</td>
<td>0.141</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Vol}} )</td>
<td>0.041***</td>
<td>0.095***</td>
<td>0.119**</td>
<td>0.135</td>
<td>0.163</td>
<td>0.164</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.067</td>
<td>0.043</td>
<td>0.033</td>
<td>0.024</td>
<td>0.020</td>
<td>0.014</td>
</tr>
<tr>
<td>LRSQ(3,1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Slp}} )</td>
<td>0.003</td>
<td>0.002</td>
<td>-0.005</td>
<td>-0.031</td>
<td>-0.059</td>
<td>-0.088</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Vol}} )</td>
<td>0.009</td>
<td>0.015</td>
<td>0.023</td>
<td>0.051</td>
<td>0.082</td>
<td>0.117</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.006</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>LRSQ(3,2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Slp}} )</td>
<td>0.001</td>
<td>0.002</td>
<td>0.007</td>
<td>0.015</td>
<td>0.015</td>
<td>0.008</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Vol}} )</td>
<td>0.013</td>
<td>0.030</td>
<td>0.045</td>
<td>0.064</td>
<td>0.076</td>
<td>0.091</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.011</td>
<td>0.009</td>
<td>0.007</td>
<td>0.005</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>LRSQ(3,3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Slp}} )</td>
<td>0.019</td>
<td>0.032</td>
<td>0.037</td>
<td>0.042</td>
<td>0.042</td>
<td>0.036</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Vol}} )</td>
<td>0.025</td>
<td>0.047</td>
<td>0.065</td>
<td>0.096</td>
<td>0.119</td>
<td>0.143</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.078</td>
<td>0.047</td>
<td>0.030</td>
<td>0.018</td>
<td>0.012</td>
<td>0.009</td>
</tr>
</tbody>
</table>
Table I.4: Conditional excess returns, zero-coupon bonds.
The table reports results from regressing nonoverlapping monthly excess zero-coupon bond returns on previous month’s term structure slope and implied volatility (including a constant). Consider the results for the 5-year maturity: The excess return is on a 5-year zero-coupon bond bootstrapped from swap rates. The term structure slope is the difference between the 5-year swap rate and 1-month LIBOR. The implied volatility is the normal implied volatility of a swaption with a 1-month option expiry and a 5-year swap maturity. Excess returns are in percent, and the term structure slopes and implied volatilities are standardized. The top panel shows results in the data, where each time series consists of 200 monthly observations from February 1997 to August 2013. t-statistics, corrected for heteroscedasticity and serial correlation up to 12 lags using the method of Newey and West (1987), are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The lower panels show results in simulated data, where each time series consists of 600,000 monthly observations (50,000 years).
J Comparison with Exponential-Affine Model

Table J.1 displays parameter estimates and their asymptotic standard errors for the \( LRSQ(3,0) \) and \( A_3(3) \) models. The \( LRSQ(3,0) \) specification is similar to its USV extensions in terms of the lower bi-diagonal structure of \( \kappa \), the effective expression for \( r_t \), the upper bound on \( r_t \) (73.2%), and the value of \( \alpha \) (5.92%). The \( A_3(3) \) model also has a lower bi-diagonal structure of \( \tilde{\kappa} \), and the first two elements of \( \gamma \) are essentially zero making \( r_t \) solely a function of \( X_{3t} \).\(^4\) Also, the second and third elements of \( \kappa \theta \) in the \( LRSQ(3,0) \) model and \( \tilde{\kappa} \tilde{\theta} \) in the \( A_3(3) \) model are approximately zero. Overall, the parameter configuration of the \( A_3(3) \) model is similar to the parameter configuration of the \( A_2(2) \) model that Kim and Singleton (2012) estimate on Japanese yield data.

The performance of the two models is illustrated in Figure J.1. Panel A1 shows time series of the 1-year, 5-year, and 10-year swap rates; Panels A2 and B2 show the fit to swap rates for the \( LRSQ(3,0) \) and \( A_3(3) \) models, respectively; and Panels A3 and B3 show time series of RMSEs for the \( LRSQ(3,0) \) and \( A_3(3) \) models, respectively. The time series of RMSEs are very similar for the two models and highly correlated with the time series of RMSEs for swap rates displayed in Figure 2, Panel A3 in the paper.

\(^4\)This allows the model to capture significant variation in longer-term interest rates during the ZLB period as \( X_{1t} \) and \( X_{2t} \) can vary without affecting \( r_t \).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>LRSQ(3.0)</th>
<th>$A_3(3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_{11}$, $\tilde{\kappa}_{11}$</td>
<td>0.0978 (0.0039)</td>
<td>0.0775 (0.0042)</td>
</tr>
<tr>
<td>$\kappa_{22}$, $\tilde{\kappa}_{22}$</td>
<td>0.4267 (0.0126)</td>
<td>0.4797 (0.0335)</td>
</tr>
<tr>
<td>$\kappa_{33}$, $\tilde{\kappa}_{33}$</td>
<td>0.6724 (0.0192)</td>
<td>0.7227 (0.0446)</td>
</tr>
<tr>
<td>$\kappa_{21}$, $\tilde{\kappa}_{21}$</td>
<td>$-0.1570$ (0.0037)</td>
<td>$-0.2258$ (0.0272)</td>
</tr>
<tr>
<td>$\kappa_{32}$, $\tilde{\kappa}_{32}$</td>
<td>$-0.4859$ (0.0125)</td>
<td>$-3.2311$ (0.3441)</td>
</tr>
<tr>
<td>$\theta_1$, $\tilde{\theta}_1$</td>
<td>0.6053 (0.0291)</td>
<td>11.9057 (1.2454)</td>
</tr>
<tr>
<td>$\theta_2$, $\tilde{\theta}_2$</td>
<td>0.2227 (0.0065)</td>
<td>5.6040 (0.5220)</td>
</tr>
<tr>
<td>$\theta_3$, $\tilde{\theta}_3$</td>
<td>0.1610 (0.0059)</td>
<td>25.0558 (1.6678)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.3468 (0.0115)</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2999 (0.0082)</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.1010 (0.0033)</td>
<td>1</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0000 (0.0000)</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0000 (0.0002)</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0032 (0.0002)</td>
<td>0</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>$-0.4513$ (0.3586)</td>
<td>0</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>$-0.7609$ (0.6950)</td>
<td>$-0.1686$ (0.1540)</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>$-3.0441$ (1.2276)</td>
<td>$-0.2641$ (1.293)</td>
</tr>
<tr>
<td>$\sigma_{rates}$</td>
<td>3.9694 (0.0501)</td>
<td>3.9737 (0.0552)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0592</td>
<td>0.0592</td>
</tr>
<tr>
<td>$supr_t$</td>
<td>0.7316</td>
<td>0.7316</td>
</tr>
<tr>
<td>$\mathcal{L} \times 10^{-4}$</td>
<td>3.0380</td>
<td>3.0428</td>
</tr>
</tbody>
</table>

Table J.1: Maximum likelihood estimates
The table reports parameter estimates with asymptotic standard errors are in parentheses. $\sigma_{rates}$ denotes the standard deviation of swap rate pricing errors measured in basis points. $\alpha$ is chosen as the smallest value that guarantees a nonnegative short rate. $supr_t$ is the upper bound on possible short rates. $\mathcal{L}$ denotes the log-likelihood values. The sample period consists of 866 weekly observations from January 29, 1997 to August 28, 2013.
Figure J.1: Data and fit
Panel A1 shows time series of the 1-year, 5-year, and 10-year swap rates (displayed as thick light-grey, thick dark-grey, and thin black lines, respectively). Panels A2 and B2 show the fit to swap rates for the \( \text{LRSQ}(3,0) \) and \( A_3(3) \) models, respectively. Panels A3 and B3 show time series of the root-mean-squared pricing errors (RMSEs) of swap rates for the \( \text{LRSQ}(3,0) \) and \( A_3(3) \) models, respectively. The units in Panels 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 866 weekly observations from January 29, 1997 to August 28, 2013.
References (Online Appendix)


