A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives

Anders B. Trolle
Copenhagen Business School

Eduardo S. Schwartz
UCLA Anderson School of Management and NBER

We develop a tractable and flexible stochastic volatility multifactor model of the term structure of interest rates. It features unspanned stochastic volatility factors, correlation between innovations to forward rates and their volatilities, quasi-analytical prices of zero-coupon bond options, and dynamics of the forward rate curve, under both the actual and risk-neutral measures, in terms of a finite-dimensional affine state vector. The model has a very good fit to an extensive panel dataset of interest rates, swaptions, and caps. In particular, the model matches the implied cap skews and the dynamics of implied volatilities. (*JEL* E43, G13)

1. Introduction

A number of stylized facts about interest rate volatility have been uncovered in the literature. First, interest rate volatility is clearly stochastic. Second, interest rate volatility contains important unspanned components. For instance, Collin-Dufresne and Goldstein (2002a); Heidari and Wu (2003); and Li and Zhao (2006) identify a number of unspanned stochastic volatility factors driving interest rate derivatives that do not affect the term structure, and Andersen and Benzoni (2005) also find unspanned factors in realized interest rate volatility. Third, changes in interest rate volatility are correlated with changes in interest rates. For instance, estimates in Andersen and Lund (1997) and Ball and Torous (1999), who both study the dynamics of the short-term interest rate, imply that relative interest rate volatility is negatively correlated with interest rates while absolute interest rate volatility is positively correlated with interest rates.[1] As

---

[1] Both papers estimate a stochastic volatility extension of the Chan et al. (1992) model given by

$$dr(t) = \kappa_1(\mu_1 - r(t))dt + \sqrt{v(t)}r(t)^{\gamma}dW_1(t),$$
$$d\log v(t) = \kappa_2(\mu_2 - \log v(t))dt + \sigma dW_2(t),$$

we discuss below, similar results are obtained from time series of implied swaption and cap volatilities. Fourth, the unconditional (realized and implied) volatility term structure exhibits a hump—see, e.g., the discussion in Dai and Singleton (2003).

In this paper, we develop a tractable and flexible multifactor model of the term structure of interest rates that is consistent with these stylized facts about interest rate volatility. The model is based on the Heath, Jarrow, and Morton (1992) (HJM, henceforth) framework. In its most general form, the model has N factors, which drive the term structure, and N additional unspanned stochastic volatility factors, which affect only interest rate derivatives. Importantly, the model allows innovations to interest rates and their volatilities to be correlated. Furthermore, the model can accommodate a wide range of shocks to the term structure including hump-shaped shocks. We derive quasi-analytical zero-coupon bond option (and therefore cap) prices based on transform techniques, while coupon bond option (and therefore swaption) prices can be obtained using well-known and accurate approximations. We show that the dynamics of the term structure under the risk-neutral probability measure can be described in terms of a finite number of state variables that jointly follow an affine diffusion process. This facilitates pricing of complex interest rate derivatives by Monte Carlo simulations. We apply the flexible "extended affine" market price of risk specification proposed by Cheredito, Filipovic, and Kimmel (2007), which implies that the state vector also follows an affine diffusion process under the actual probability measure and facilitates the application of standard econometric techniques.

We estimate the model for $N = 1$, 2, and 3 using an extensive panel dataset consisting of 7 years (plus 1.5 years of additional data used for out-of-sample analysis) of weekly observations of LIBOR and swap rates, at-the-money-forward (ATMF, henceforth) swaptions, ATMF caps, and for the second half of the sample, non-ATMF caps (i.e., cap skews). To our knowledge, this is the most extensive dataset, in terms of the range of instruments included, that has been used in the empirical term structure literature to date. The estimation procedure is quasi maximum likelihood in conjunction with the extended Kalman filter.

The empirical part of the paper contains a number of contributions. First, we show that for $N = 3$, the model has a very good fit to both interest rates and interest rate derivatives. This is consistent with principal component analyses that show that three factors are necessary to capture the variation in the term structure (see, e.g., Litterman and Scheinkman, 1991) and, as discussed above, that a number of additional unspanned stochastic volatility factors are needed to explain the variation in interest rate derivatives fully. This conclusion also holds true in the out-of-sample period.

Second, we address the relative valuation of swaptions and caps by reestimating the $N = 3$ model separately on swaptions and caps, and pricing caps

---

where the correlation between $W_1(t)$ and $W_2(t)$ is set to zero. The short-term interest rate and its volatility are correlated through the term $r(t)^\gamma$. Andersen and Lund (1997) estimate $\gamma = 0.544$ and Ball and Torous (1999) estimate $\gamma = 0.754$ implying the dynamics stated in the text.

and swaptions out of sample. We find that, according to our model, swaptions were mostly undervalued relative to caps during the first 2.5 years of the sample. However, since then swaption and cap prices appear largely consistent with each other.

Third, we stress the importance of allowing innovations to interest rates and their volatilities to be correlated. In the cross-sectional dimension of the data, we observe downward sloping cap skews in terms of lognormal implied volatilities with low-strike, in-the-money caps trading at higher lognormal implied volatilities than high-strike, out-of-the-money caps. In the time-series dimension of the data, we observe that changes in lognormal implied volatilities of both swaptions and caps are moderately negatively correlated with changes in the underlying forward rates, while changes in normal implied volatilities are moderately positively correlated with changes in the underlying forward rates.[2,3] In our model, both the steepness of the implied cap skews and the dynamics of implied volatilities depend critically on the correlation parameters, and the model is able to match both features of the data accurately. In other words, our model provides a consistent explanation of why and how implied volatilities vary across moneyness and time.

Fourth, we test the $N = 3$ models estimated separately on swaptions and caps against a range of nested models. The fit to both interest rates and interest rate derivatives becomes progressively worse as more of the term structure factors are restricted to generate exponentially declining, rather than more flexible and possibly hump-shaped, innovations to the forward rate curve and as the number of unspanned stochastic volatility factors is reduced. Furthermore, the ability to fit the cap skew deteriorates significantly if innovations to interest rates and their volatilities are assumed uncorrelated. This shows that all the major features of our model are necessary to provide an adequate fit to the entire dataset.

Our model is related to the stochastic volatility LIBOR market models of Han (2007) and Jarrow, Li, and Zhao (2007). Han (2007) estimates his model on swaption data, while Jarrow, Li, and Zhao (2007) estimate their model on cap skew data. In their models, conditional on the volatility state variables, forward LIBOR rates are lognormally distributed, and forward swap rates are approximately lognormally distributed (under the appropriate forward measures). In

---

[2] In this paper, the term "lognormal implied volatility" is the volatility parameter that, plugged into the lognormal (or Black, 1976) pricing formula, matches a given price. The term "normal implied volatility" is the volatility parameter that, plugged into the normal pricing formula, matches a given price. For ATMF swaptions or caplets, the relation between the two is approximately given by $\sigma_N = \sigma_{LN} F(t, T)$, where σ_N is the normal implied volatility, σ_{LN} is the lognormal implied volatility and $F(t, T)$ is the underlying forward rate.

[3] The average correlation between weekly changes in lognormal (normal) implied volatilities and weekly changes in the underlying forward rates is -0.354 (0.349) for the 42 ATMF swaptions in the dataset and -0.331 (0.347) for the seven ATMF caps in the dataset. Surprisingly, Chen and Scott (2001) report that the correlation between changes in the lognormal implied volatilities from options on short-term Eurodollar futures and changes in the underlying futures rates is only -0.07. This may have to do with their using a different sample period from ours (they consider an earlier period from March 1985 to December 2000) and the fact that they consider the very short end of the yield curve, which is highly affected by Fed behavior (see, e.g., Piazzesi, 2005).

contrast, in our model, conditional on the volatility state variables, forward LIBOR and swap rates are approximately normally distributed (under the appropriate forward measures). More importantly, to make their models tractable, they impose zero correlation between innovations to forward LIBOR rates and their volatilities. The zero-correlation assumption implies that the forward LIBOR rate distributions have fatter tails than the lognormal distribution, and their models predict implied volatility smiles rather than the implied volatility skews observed in the data.[4] To match the implied volatility skews, Jarrow, Li, and Zhao (2007) add jumps to the forward rate processes and estimate large negative mean jump sizes (under the forward measures).

There are two issues with the zero-correlation constraint in their models, however. First, their models are not able to match the dynamics of implied volatilities across time as they imply that changes in lognormal implied volatilities are approximately uncorrelated with changes in the underlying forward rates. Indeed, Jarrow, Li, and Zhao (2007) report that the state variable that drives most of the stochastic volatility is strongly negatively correlated with interest rates despite the zero-correlation constraint in their model. Second, we show that allowing for correlation between innovations to forward rates and their volatilities can account for much of the implied volatility skew. By ignoring this aspect, Jarrow, Li, and Zhao (2007) may overstate the importance of jumps for pricing non-ATMF caps.

It seems logical, then, to extend the stochastic volatility LIBOR market model (possibly with jumps) to nonzero correlation between innovations to forward LIBOR rates and volatility. Unfortunately, such a model is intractable.[5] The ease with which we can incorporate nonzero correlation is one reason we prefer to work with instantaneous forward rates within the HJM framework. Another reason is our ability to obtain a finite-dimensional affine model of the evolution of the forward rate curve.[6]

---

[4] In contrast, in our model, a zero correlation assumption would imply that the forward LIBOR rate distributions have fatter tails than the normal distribution and the model would predict very steep lognormal implied skews—steeper than observed in the data.

[5] The reason why nonzero correlation undermines the tractability of a stochastic volatility LIBOR market model is that the dynamics of the volatility process becomes dependent on forward rates under the forward measure. See Wu and Zhang (2005) for more on this issue and the approximations necessary to retain analytical tractability, even with nonzero correlation. Andersen and Brotherton-Ratciffe (2005) develop a LIBOR market model with unspannned stochastic volatility factors in which the forward rates enter the diffusion terms of the forward rate processes in a flexible way that allows forward rates and their volatilities to be correlated. Pricing of caps and swaptions relies on a number of fairly involved approximations, and they make no attempt to test their model on a panel dataset of interest rate derivatives.

[6] In LIBOR market models, it is typically not possible to obtain a finite-dimensional Markov model for the evolution of the forward rate curve. Apart from making pricing by simulations more complicated, it also prohibits estimating the model simultaneously on interest rates and derivatives by standard approaches. Instead, Han (2007) and Jarrow, Li, and Zhao (2007) apply a two-step estimation approach in which, first, the loadings on the term structure factors are obtained as the eigenvectors from a factor analysis of the historical covariance matrix of forward rates, and second, the parameters of the volatility processes (and possibly the jumps) are estimated from interest rate derivatives. Estimating a model simultaneously on interest rates and derivatives, as we do in this paper, gives additional flexibility in terms of fitting the data.

Our model is also related to Casassus, Collin-Dufresne, and Goldstein (2005), who develop a stochastic volatility Hull and White (1990) model, which is a special case of our model. Using implied cap skew data on a single date, they also document the importance of allowing for nonzero correlation between innovations to forward rates and volatility.

Other papers that use interest rate derivatives for estimating dynamic term structure models include Umantsev (2001), who uses swaptions, Bikbov and Chernov (2004), who use options on Eurodollar futures, and Almeida, Graveline, and Joslin (2006), who use caps. These papers estimate traditional three-factor affine models that do not have sufficient flexibility to match the extensive dataset used in this paper. Furthermore, in these models it is very difficult to generate unspanned stochastic volatility that arises naturally within the HJM framework.[7]

The paper is structured as follows. Section 2 describes our general stochastic volatility term structure model. Section 3 discusses the data and the estimation procedure. Section 4 contains the estimation results. Section 5 concludes.

2. A General Stochastic Volatility Term Structure Model

2.1 The model under the risk-neutral measure

Let $f(t, T)$ denote the time-t instantaneous forward interest rate for risk-free borrowing and lending at time T. We model the forward rate dynamics as

$$df(t, T) = \mu_f(t, T)dt + \sum_{i=1}^{N} \sigma_{f,i}(t, T)\sqrt{v_i(t)}dW_i^Q(t), \qquad (1)$$

$$dv_i(t) = \kappa_i(\theta_i - v_i(t))dt + \sigma_i\sqrt{v_i(t)}\left(\rho_i dW_i^Q(t) + \sqrt{1 - \rho_i^2}dZ_i^Q(t)\right), \quad (2)$$

$i = 1, \ldots, N$, where $W_i^Q(t)$ and $Z_i^Q(t)$ denote independent standard Wiener processes under the risk-neutral measure Q. The model extends traditional HJM models by incorporating stochastic volatility. The forward rate curve is driven by N factors. Forward rate volatilities, and hence interest rate derivatives, are driven by $N \times 2$ factors, except if $\rho_i = -1$ or $\rho_i = 1$ for some i. Innovations to forward rates and their volatilities are correlated, except if $\rho_i = 0$ for all i. For $N = 1$, the model can be seen as the fixed-income counterpart to the Heston (1993) model, which has been used extensively in the equity derivatives literature.

---

[7] See Collin-Dufresne and Goldstein (2002a) for the parameter restrictions that are necessary in order for traditional three-factor affine models to exhibit unspanned stochastic volatility. They also show that traditional affine models with two factors or less, such as the Longstaff and Schwartz (1992) model, cannot exhibit unspanned stochastic volatility.

Heath, Jarrow, and Morton (1992) show that absence of arbitrage implies that the drift term in Equation (1) is given by

$$\mu_f(t, T) = \sum_{i=1}^{N} v_i(t)\sigma_{f,i}(t, T) \int_t^T \sigma_{f,i}(t, u)du. \quad (3)$$

Hence, the dynamics of $f(t, T)$ under Q are completely determined by the initial forward rate curve, the forward rate volatility functions, $\sigma_{f,i}(t, T)$, and the volatility state variables, $v_i(t)$.

For a general specification of $\sigma_{f,i}(t, T)$, the dynamics of the forward rate curve will be path-dependent, which significantly complicates derivatives pricing and the application of standard econometric techniques. A branch of the term structure literature has investigated under which conditions HJM models are Markovian with respect to a finite number of state variables.[8] Applying these results to our setting, it can be shown that a sufficient condition for the dynamics of the forward rate curve to be represented by a finite-dimensional Markov process and for the volatility structure to be time-homogeneous is that $\sigma_{f,i}(t, T) = p_n(T - t)e^{-\gamma_i(T-t)}$, where $p_n(\tau)$ is an n-order polynomial in τ. To keep the model flexible yet tractable, we set set $n = 1$ such that

$$\sigma_{f,i}(t, T) = (\alpha_{0,i} + \alpha_{1,i}(T - t))e^{-\gamma_i(T-t)}. \quad (4)$$

This specification allows for a wide range of shocks to the forward rate curve. In particular, it allows for hump-shaped shocks that turn out to be essential to match interest rate derivatives.[9] Furthermore, the specification nests a number of interesting special cases. With $N = 1$ and $\alpha_{1,1} = 0$, we get the stochastic volatility version of the Hull and White (1990) model analyzed by Casassus, Collin-Dufresne, and Goldstein (2005). When also $\gamma_1 = 0$, we obtain a stochastic volatility version of the continuous-time Ho and Lee (1986) model.

The following proposition shows the Markov representation of the model:

Proposition 1. *The time-t instantaneous forward interest rate for risk-free borrowing and lending at time T, $f(t, T)$, is given by*

$$f(t, T) = f(0, T) + \sum_{i=1}^{N} \mathcal{B}_{x_i}(T - t)x_i(t) + \sum_{i=1}^{N} \sum_{j=1}^{6} \mathcal{B}_{\phi_{j,i}}(T - t)\phi_{j,i}(t), \quad (5)$$

---

[8] See, e.g., Ritchken and Sankarasubramaniam (1995); Bhar and Chiarella (1997); Inui and Kijima (1998); de Jong and Santa-Clara (1999); Ritchken and Chuang (1999); and Chiarella and Kwon (2003).

[9] Note that $\alpha_{0,i}$, $\alpha_{1,i}$, θ_i, and σ_i are not simultaneously identified; see, e.g., the discussion of invariant affine transformations in Dai and Singleton (2000). In our empirical analysis, we normalize θ_i to 1 to achieve identification.

where

$$\mathcal{B}_{x_i}(\tau) = (\alpha_{0i} + \alpha_{1i}\tau)e^{-\gamma_i\tau}, \tag{6}$$

$$\mathcal{B}_{\phi_{1,i}}(\tau) = \alpha_{1i}e^{-\gamma_i\tau}, \tag{7}$$

$$\mathcal{B}_{\phi_{2,i}}(\tau) = \frac{\alpha_{1i}}{\gamma_i}\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right)(\alpha_{0i} + \alpha_{1i}\tau)e^{-\gamma_i\tau}, \tag{8}$$

$$\mathcal{B}_{\phi_{3,i}}(\tau) = -\left(\frac{\alpha_{0i}\alpha_{1i}}{\gamma_i}\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right) + \frac{\alpha_{1i}}{\gamma_i}\left(\frac{\alpha_{1i}}{\gamma_i} + 2\alpha_{0i}\right)\tau + \frac{\alpha_{1i}^2}{\gamma_i}\tau^2\right)e^{-2\gamma_i\tau}, \tag{9}$$

$$\mathcal{B}_{\phi_{4,i}}(\tau) = \frac{\alpha_{1i}^2}{\gamma_i}\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right)e^{-\gamma_i\tau}, \tag{10}$$

$$\mathcal{B}_{\phi_{5,i}}(\tau) = -\frac{\alpha_{1i}}{\gamma_i}\left(\frac{\alpha_{1i}}{\gamma_i} + 2\alpha_{0i} + 2\alpha_{1i}\tau\right)e^{-2\gamma_i\tau}, \tag{11}$$

$$\mathcal{B}_{\phi_{6,i}}(\tau) = -\frac{\alpha_{1i}^2}{\gamma_i}e^{-2\gamma_i\tau}, \tag{12}$$

and the state variables evolve according to

$$dx_i(t) = -\gamma_i x_i(t)dt + \sqrt{v_i(t)}dW_i^Q(t), \tag{13}$$

$$d\phi_{1,i}(t) = (x_i(t) - \gamma_i\phi_{1,i}(t))dt, \tag{14}$$

$$d\phi_{2,i}(t) = (v_i(t) - \gamma_i\phi_{2,i}(t))dt, \tag{15}$$

$$d\phi_{3,i}(t) = (v_i(t) - 2\gamma_i\phi_{3,i}(t))dt, \tag{16}$$

$$d\phi_{4,i}(t) = (\phi_{2,i}(t) - \gamma_i\phi_{4,i}(t))dt, \tag{17}$$

$$d\phi_{5,i}(t) = (\phi_{3,i}(t) - 2\gamma_i\phi_{5,i}(t))dt, \tag{18}$$

$$d\phi_{6,i}(t) = (2\phi_{5,i}(t) - 2\gamma_i\phi_{6,i}(t))dt, \tag{19}$$

subject to $x_i(0) = \phi_{1,i}(0) = \cdots = \phi_{6,i}(0) = 0.$

Proof.   See Appendix A. ∎

Note that forward rates do not depend directly on the volatility state variables. The dynamics of the forward rate curve are given in terms of $N \times 8$ state variables that jointly follow an affine diffusion process. There are no stochastic terms in the $\phi_{1,i}(t), \ldots, \phi_{6,i}(t)$ processes, which are "auxiliary," locally deterministic, state variables that reflect the path information of $x_i(t)$ and $v_i(t)$. By augmenting the state space with these variables, the model becomes Markovian. The model falls within the affine class of dynamic term structure models of Duffie and Kan (1996) and inherits all the nice analytical features of that class. The model is time-inhomogeneous, as the dynamics of the forward rate curve depends on the initial term structure. In Section 3, we reduce the model to its time-homogeneous counterpart for the purpose of econometric estimation.

2.2 Prices of zero-coupon bonds and bond options

The time-t price of a zero-coupon bond maturing at time T, $P(t, T)$, is given by

$$
\begin{aligned}
P(t, T) &\equiv \exp\left\{-\int_t^T f(t, u)du\right\} \\
&= \frac{P(0, T)}{P(0, t)} \exp\left\{\sum_{i=1}^N B_{x_i}(T - t)x_i(t) + \sum_{i=1}^N \sum_{j=1}^6 B_{\phi_{j,i}}(T - t)\phi_{j,i}(t)\right\},
\end{aligned}
$$

(20)

where

$$
B_{x_i}(\tau) = \frac{\alpha_{1i}}{\gamma_i}\left(\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right)(e^{-\gamma_i \tau} - 1) + \tau e^{-\gamma_i \tau}\right),
$$

(21)

$$
B_{\phi_{1,i}}(\tau) = \frac{\alpha_{1i}}{\gamma_i}(e^{-\gamma_i \tau} - 1),
$$

(22)

$$
B_{\phi_{2,i}}(\tau) = \left(\frac{\alpha_{1i}}{\gamma_i}\right)^2 \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right)\left(\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right)(e^{-\gamma_i \tau} - 1) + \tau e^{-\gamma_i \tau}\right),
$$

(23)

$$
\begin{aligned}
B_{\phi_{3,i}}(\tau) = &-\frac{\alpha_{1i}}{\gamma_i^2}\left(\left(\frac{\alpha_{1i}}{2\gamma_i^2} + \frac{\alpha_{0i}}{\gamma_i} + \frac{\alpha_{0i}^2}{2\alpha_{1i}}\right)(e^{-2\gamma_i \tau} - 1) \right. \\
&\left. + \left(\frac{\alpha_{1i}}{\gamma_i} + \alpha_{0i}\right)\tau e^{-2\gamma_i \tau} + \frac{\alpha_{1i}}{2}\tau^2 e^{-2\gamma_i \tau}\right),
\end{aligned}
$$

(24)

$$
B_{\phi_{4,i}}(\tau) = \left(\frac{\alpha_{1i}}{\gamma_i}\right)^2 \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}}\right)(e^{-\gamma_i \tau} - 1),
$$

(25)

$$
B_{\phi_{5,i}}(\tau) = -\frac{\alpha_{1i}}{\gamma_i^2}\left(\left(\frac{\alpha_{1i}}{\gamma_i} + \alpha_{0i}\right)(e^{-2\gamma_i \tau} - 1) + \alpha_{1i}\tau e^{-2\gamma_i \tau}\right),
$$

(26)

$$
B_{\phi_{6,i}}(\tau) = -\frac{1}{2}\left(\frac{\alpha_{1i}}{\gamma_i}\right)^2 (e^{-2\gamma_i \tau} - 1).
$$

(27)

It follows that the dynamics of $P(t, T)$ is given by

$$
\frac{dP(t, T)}{P(t, T)} = r(t)dt + \sum_{i=1}^N B_{x_i}(T - t)\sqrt{v_i(t)}dW_i^Q(t).
$$

(28)

To price options on zero-coupon bonds, we follow Collin-Dufresne and Goldstein (2003), who extend the analysis in Duffie, Pan, and Singleton (2000) to HJM models, and introduce the transform

$$
\psi(u, t, T_0, T_1) = E_t^Q\left[e^{-\int_t^{T_0} r_s ds} e^{u \log(P(T_0, T_1))}\right].
$$

(29)

This transform has an exponentially affine solution as demonstrated in the following proposition:

Proposition 2. *The transform in (29) is given by*

$$\psi(u, t, T_0, T_1) = e^{M(T_0-t)+\sum_{i=1}^{N} N_i(T_0-t)v_i(t)+u\log(P(t,T_1))+(1-u)\log(P(t,T_0))}, \quad (30)$$

where $M(\tau)$ and $N_i(\tau)$ solve the following system of ODEs:

$$\frac{dM(\tau)}{d\tau} = \sum_{i=1}^{N} N_i(\tau)\kappa_i\theta_i, \quad (31)$$

$$\begin{aligned}
\frac{dN_i(\tau)}{d\tau} &= N_i(\tau)(-\kappa_i + \sigma_i\rho_i(uB_{x_i}(T_1 - T_0 + \tau) + (1-u)B_{x_i}(\tau))) \\
&\quad + \tfrac{1}{2}N_i(\tau)^2\sigma_i^2 + \tfrac{1}{2}(u^2 - u)B_{x_i}(T_1 - T_0 + \tau)^2 \\
&\quad + \tfrac{1}{2}((1-u)^2 - (1-u))B_{x_i}(\tau)^2 \\
&\quad + u(1-u)B_{x_i}(T_1 - T_0 + \tau)B_{x_i}(\tau), \quad (32)
\end{aligned}$$

subject to the boundary conditions $M(0) = 0$ and $N_i(0) = 0$.

Proof. See Appendix A. ∎

As in Duffie, Pan, and Singleton (2000) and Collin-Dufresne and Goldstein (2003), we can now price options on zero-coupon bonds by applying the Fourier inversion theorem.

Proposition 3. *The time-t price of a European put option expiring at time T_0 with strike K on a zero-coupon bond maturing at time T_1, $\mathcal{P}(t, T_0, T_1, K)$, is given by*

$$\mathcal{P}(t, T_0, T_1, K) = K G_{0,1}(\log(K)) - G_{1,1}(\log(K)), \quad (33)$$

where $G_{a,b}(y)$ is defined as

$$G_{a,b}(y) = \frac{\psi(a, t, T_0, T_1)}{2} - \frac{1}{\pi}\int_0^\infty \frac{\text{Im}[\psi(a + \mathtt{i}ub, t, T_0, T_1)e^{-\mathtt{i}uy}]}{u}du, \quad (34)$$

where $\mathtt{i} = \sqrt{-1}$.

Proof. See Appendix A. ∎

For estimation, we will use LIBOR rates, swap rates, caps, and swaptions. LIBOR and swap rates are straightforward to compute from the zero-coupon curve. A cap is a portfolio of caplets. A caplet is a call option on a LIBOR rate but can also be valued as a (scaled) European put option on a zero-coupon bond and can therefore be priced using Proposition 3. A payer swaption is a call option

on a swap rate but can also be valued as a European put option on a coupon bond. No analytical expressions exist for European coupon bond options in the general affine framework, but a number of accurate approximations have been developed. We apply the stochastic duration approach developed by Wei (1997) for one-factor models and extended to multifactor models by Munk (1999). This approximation is fast and has been shown to be accurate for ATMF options, which is what we use for estimation; see Munk (1999) and Singleton and Umantsev (2002).[10] The idea of the stochastic duration approach is to approximate a European option on a coupon bond with a (scaled) European option on a zero-coupon bond with maturity equal to the stochastic duration of the coupon bond. Therefore, swaptions can also be priced using Proposition 3. Appendix B contains the pricing formulas for LIBOR rates, swap rates, caps, and swaptions.

2.3 Implications for implied volatilities

Our model is expressed in terms of instantaneous forward rates. In contrast, LIBOR market models (Miltersen, Sandmann, and Sondermann, 1997; and Brace, Gatarek, and Musiela, 1997) are expressed in terms of forward LIBOR rates, while swap market models (Jamshidian, 1997) are expressed in terms of forward swap rates. In this section, we relate our model to these competing frameworks popular in the financial industry. We also obtain very intuitive formulas for the ATMF implied volatilities for swaptions and caplets in our model.[11]

Applying Ito's Lemma to the time-u forward swap rate for the period T_m to T_n [see Equation (76) in Appendix B] and switching to the forward swap measure under which forward swap rates are martingales (see Jamshidian, 1997), we obtain

$$dS(u, T_m, T_n) = \sum_{i=1}^{N} \left(\sum_{j=m}^{n} \zeta_j(u) B_{x_i}(T_j - u) \right) \sqrt{v_i(u)} dW_i^{Q^{T_m, T_n}}(u), \quad (35)$$

where $\zeta_m(u) = \frac{P(u, T_m)}{PVBP(u)}$, $\zeta_j(u) = -vS(u, T_m, T_n)\frac{P(u, T_j)}{PVBP(u)}$ for $j = m + 1, \dots, n - 1$, $\zeta_n(u) = -(1 + vS(u, T_m, T_n))\frac{P(u, T_n)}{PVBP(u)}$ and $PVBP(u) = v \sum_{j=m+1}^{n} P(u, T_j)$. Furthermore, the dynamics of $v_i(u)$ under the forward swap measure

---

[10] Other approximation schemes have been developed by Collin-Dufresne and Goldstein (2002b); Singleton and Umantsev (2002); and Schrager and Pelsser (2006). However, these tend to be slower than the stochastic duration approach and hence not well suited for this paper, in which a very large number of swaption prices needs to be computed for each evaluation of the likelihood function.

[11] To keep the discussion brief, we will focus on the dynamics of forward swap rates and ATMF swaption implied volatilities. However, since a forward LIBOR rate can be seen as a particular forward swap rate, the analysis also applies to the dynamics of forward LIBOR rates and ATMF caplet implied volatilities.

are given by

$$
dv_i(u) = \left(\kappa_i(\theta_i - v_i(u)) + v_i(u)\sigma_i\rho_i\nu \sum_{j=m+1}^{n} \xi_j(u)B_{x_i}(T_j - u) \right) du
$$
$$
+ \sigma_i\sqrt{v_i(u)}\left(\rho_i dW_i^{Q^{T_m,T_n}}(u) + \sqrt{1 - \rho_i^2}dZ_i^{Q^{T_m,T_n}}(u)\right), \tag{36}
$$

where $\xi_j(u) = \frac{P(u,T_j)}{PVBP(u)}$. $W_i^{Q^{T_m,T_n}}(u)$ and $Z_i^{Q^{T_m,T_n}}(u)$ denote independent standard Wiener processes under the forward swap measure Q^{T_m,T_n}.

While instantaneous forward rates are normally distributed conditional on the volatility state variables, the same does not hold for forward swap rates, since the $\zeta_j(u)$ terms are stochastic. Also, the process of $v_i(u)$ is nonaffine under the forward swap measure due to the stochastic $\xi_j(u)$ terms. However, we can obtain an approximate and affine expression for the dynamics of the forward swap rate by replacing $\zeta_j(u)$ and $\xi_j(u)$ with their time-t expected values, which are simply their time-t values since these terms are martingales under the forward swap measure.[12] This implies that, conditional on the volatility state variables, forward swap (and LIBOR) rates are approximately normally distributed in our model. This is in contrast to the LIBOR and swap market models where forward swap (and LIBOR) rates are typically (either approximately or exactly) lognormally distributed.[13]

We can make a second approximation by replacing $v_i(u)$ in Equation (35) with its time-t expected value. In this case, given time-t information, $S(T_m, T_n)$ is normally distributed

$$
S(T_m, T_n) \sim \mathcal{N}(S(t, T_m, T_n), \sigma_N(t, T_m, T_n)\sqrt{T_m - t}), \tag{37}
$$

where

$$
\sigma_N(t, T_m, T_n)
$$
$$
= \left(\frac{1}{T_m - t} \int_t^{T_m} \sum_{i=1}^{N} \left(\sum_{j=m}^{n} \zeta_j(t)B_{x_i}(T_j - u) \right)^2 E_t^{Q^{T_m,T_n}}[v_i(u)]du \right)^{1/2}. \tag{38}
$$

Then, an approximate price of a $(T_m - t)$–into–$(T_n - T_m)$ swaption (i.e., the time-t price of an option expiring at T_m on a swap for the period T_m to T_n) can be obtained by inserting Equation (38) in the normal swaption pricing formula.

---

[12] This is because *PVBP(u)*, which is the numeraire associated with the forward swap measure, appears in the denominators of these terms. A similar approach is followed by Schrager and Pelsser (2006) in a general affine model. They argue that the approximation is very accurate since $\zeta_j(u)$ and $\xi_j(u)$ typically have low variances.

[13] The fact that forward rates are conditionally normally distributed implies that forward rates may become negative. However, for typical parameter estimates reported in Section 4, the probability of forward rates taking negative values under Q is virtually zero. The probability is generally higher, although still small, under P.

Monte Carlo evidence (not reported) shows this to be reasonably accurate for ATMF swaptions.[14] Therefore, we can view $\sigma_N(t, T_m, T_n)$ as a reasonably accurate expression for the normal implied ATMF swaption volatility in our model. The corresponding lognormal implied ATMF swaption volatility is approximately given by

$$\sigma_{LN}(t, T_m, T_n) = \frac{\sigma_N(t, T_m, T_n)}{S(t, T_m, T_n)}. \tag{39}$$

These expressions yield several insights. First, Equations (38) and (39) directly link the volatility state variables in our model to the ATMF normal and lognormal implied volatilities. A positive $v_i(t)$-shock naturally increases normal and lognormal implied volatilities. However, since $\sigma_N(t, T_m, T_n)$ equals the square root of the average expected instantaneous variance of the forward swap rate over the life of the swaptions[15] and since a $v_i(t)$-shock is expected to die out over time, the effect on implied volatilities will tend to decrease with the length of the option. Other things being equal, the effect on longer-term options will be larger for the more persistent volatility state variables.

Second, shocks to the term structure have only an indirect effect on $\sigma_N(t, T_m, T_n)$ through the $\zeta_j(t)$ and $\xi_j(t)$ terms. This effect is small for reasonable parameter values. In contrast, shocks to the term structure have a direct effect on $\sigma_{LN}(t, T_m, T_n)$ through the underlying forward rate. Therefore, in our model the normal implied volatility surface is driven almost exclusively by variations in the volatility state variables, while the lognormal implied volatility surface is driven by variations in both the volatility state variables and the term structure.

Third, and related, without correlation between innovations to the volatility state variables and the term structure, the model implies that changes in normal implied volatilities are approximately uncorrelated with changes in the underlying forward rates, while changes in lognormal implied volatilities are quite strongly negatively correlated with changes in the underlying forward rates. However, with positive correlation parameters, the model implies positive (less negative) correlations between normal (lognormal) implied volatility changes and forward rate changes, more in line with what we see in the data.

2.4 Market price of risk specifications
For estimation, we also need the dynamics of the state vector under the actual measure P, which are obtained by specifying the market prices of risk, $\Lambda_{W,i}$

[14] For $N = 3$ and typical parameter estimates reported in Section 4, the pricing errors range from -2% to 3% of the true price depending on the swaption and the values of the state variables. Note that this approach to pricing swaptions is extremely fast, requiring only a single numerical integration. Therefore, we use it in the initial stages of the estimation procedure to obtain a set of parameter estimates that is subsequently refined by applying the more accurate stochastic duration approach described in Appendix B.

[15] Where the expectation is taken under the forward swap measure.

and $\Lambda_{Z,i}$, that link the Wiener processes under Q and P through

$$dW_i^P(t) = dW_i^Q(t) - \Lambda_{W,i}(t)dt, \tag{40}$$

$$dZ_i^P(t) = dZ_i^Q(t) - \Lambda_{Z,i}(t)dt. \tag{41}$$

We apply the "extended affine" market price of risk specification suggested by Cheredito, Filipovic, and Kimmel (2007) and Collin-Dufresne, Goldstein, and Jones (2003). This is the most flexible market price of risk specification that preserves the affine structure of the state vector under the change of measure. In our setting, the "extended affine" specification is given by

$$\Lambda_{W,i}(t) = \frac{\lambda_{W,i0} + \lambda_{W,ix}x_i(t) + \lambda_{W,iv}v_i(t)}{\sqrt{v_i(t)}}, \tag{42}$$

$$\Lambda_{Z,i}(t) = \frac{1}{\sqrt{1 - \rho_i^2}} \frac{\lambda_{Z,i0} + \lambda_{Z,iv}v_i(t) - \rho_i(\lambda_{W,i0} + \lambda_{W,ix}x_i(t) + \lambda_{W,iv}v_i(t))}{\sqrt{v_i(t)}}, \tag{43}$$

which implies that the dynamics of $x_i(t)$ and $v_i(t)$ under P are given by

$$dx_i(t) = \left(\eta_i^P + \kappa_{x,i}^P x_i(t) + \kappa_{xv,i}^P v_i(t)\right) dt + \sqrt{v_i(t)}dW_i^P(t), \tag{44}$$

$$dv_i(t) = \kappa_i^P\left(\theta_i^P - v_i(t)\right)dt + \sigma_i\sqrt{v_i(t)}\left(\rho_i dW_i^P(t) + \sqrt{1 - \rho_i^2}dZ_i^P(t)\right), \tag{45}$$

where $\eta_i^P = \lambda_{W,i0}$, $\kappa_{x,i}^P = (\lambda_{W,ix} - \gamma_i)$, $\kappa_{xv,i}^P = \lambda_{W,iv}$, $\kappa_i^P = \kappa_i - \sigma_i\lambda_{Z,iv}$ and $\theta_i^P = \frac{\kappa_i\theta_i + \sigma_i\lambda_{Z,i0}}{\kappa_i^P}$. Obviously, the dynamics of $\phi_{1,i}(t), \dots, \phi_{6,i}(t)$ do not change since these contain no stochastic terms.

The traditional "completely affine" specification (see, e.g., Dai and Singleton, 2000) is obtained by setting $\lambda_{W,i0} = \lambda_{W,ix} = \lambda_{Z,i0} = 0$, while the "essentially affine" specification (see, e.g., Dai and Singleton, 2002; and Duffee, 2002) is obtained by setting $\lambda_{Z,i0} = 0$.[16] In both cases, we have that $\theta_i^P = \frac{\kappa_i\theta_i}{\kappa_i^P}$. The advantage of the "extended affine" specification is that one can adjust the mean reversion speed and the long-run level of the volatility processes independently of each other when changing measure. In contrast, with the "completely affine" and "essentially affine" specifications, adjusting the mean reversion speed necessarily changes the long-run level by a given amount.

The "extended affine" specification is only valid provided that $v_i(t)$ does not attain its boundary value of zero under both Q and P. Therefore, we have to

---

[16] Strictly speaking, in our setting, the "essentially affine" specification coincides with the "completely affine" specification. However, we could allow $dx_i(t) = \cdots dt + \sqrt{\epsilon + v_i(t)}dW_i^Q(t)$, in which case $\Lambda_{W,i}(t) = \frac{\lambda_{W,i0} + \lambda_{W,ix}x_i(t) + \lambda_{W,iv}v_i(t)}{\sqrt{\epsilon + v_i(t)}}$ and the statement in the text would be exactly correct. See Cheredito, Filipovic, and Kimmel (2007) for more on this issue.

impose the following boundary nonattainment conditions:[17]

$$2\kappa_i\theta_i \geq \sigma_i^2, \qquad (46)$$

$$2\kappa_i^P\theta_i^P \geq \sigma_i^2. \qquad (47)$$

3. Estimation Approach

3.1 Data

Our dataset consists of weekly observations of LIBOR/swap term structures and lognormal implied ATMF swaption and cap volatilities from August 21, 1998 (i.e., just prior to the LTCM crisis) to January 26, 2007. From January 4, 2002 to January 26, 2007, we also have weekly observations on the lognormal implied cap skews.[18] All observations are closing midquotes on Fridays and are obtained from Bloomberg.[19]

The LIBOR/swap term structures consist of LIBOR rates with maturities of 3, 6, and 9 months and swap rates with maturities 1, 2, 3, 5, 7, 10, and 15 years. The term structure data are displayed in Figure 1.

The swaptions have underlying swap maturities of 1, 2, 3, 5, 7, and 10 years (called "tenors") and option maturities of 1 month, 3 months, 6 months, 1 year, 2 years, 3 years, and 5 years—i.e., a total of 42 swaptions. The strikes on the ATMF swaptions are simply the forward rates on the underlying swaps. Figure 2 displays the swaption data.

The caps have length 1, 2, 3, 4, 5, 7, and 10 years. The strikes on the ATMF caps are the swap rates on the swaps with payments that correspond to those of the caps. The skew data consists of implied volatilities on caps with fixed strikes of 1.5%, 2.0%, 3.0%, 4.0%, 5.0%, 6.0%, and 7.0%. We define "moneyness" of a given cap as the ratio between its strike and the strike on the ATMF cap of the same length. Therefore, those caps with moneyness larger than one are out-of-the-money (OTM), while those with moneyness less than one are in-the-money (ITM). Rather than work with caps with fixed strikes (and time-varying moneyness), we will work with caps with fixed moneyness (and time-varying strikes) between 0.80 and 1.20. The strike on a cap with a given moneyness is obtained by cubic-spline interpolation. Figure 3 displays the ATMF cap data, while Figure 4 displays the cap skew data. The missing data in the time series of skews for the 1- and 2-year caps is due to the fact that very low interest rates have made a full skew unavailable in some periods

---

[17] Intuitively, if $v_i(t)$ were zero, we would have an infinite market price of risk despite zero volatility, representing an arbitrage opportunity. The boundary nonattainment conditions ensure that the market prices of risk stay finite, although they can become arbitrarily large. The boundary nonattainment conditions must be satisfied under both P and Q for the measures to the equivalent. See Cheredito, Filipovic, and Kimmel (2007) for a further discussion.

[18] Presently, information on implied swaption skews is not available through standard data sources.

[19] Note that we are implicitly assuming homogeneous credit quality across the LIBOR, swap, swaption, and cap markets since all cash-flows are discounted using the same discount factors.
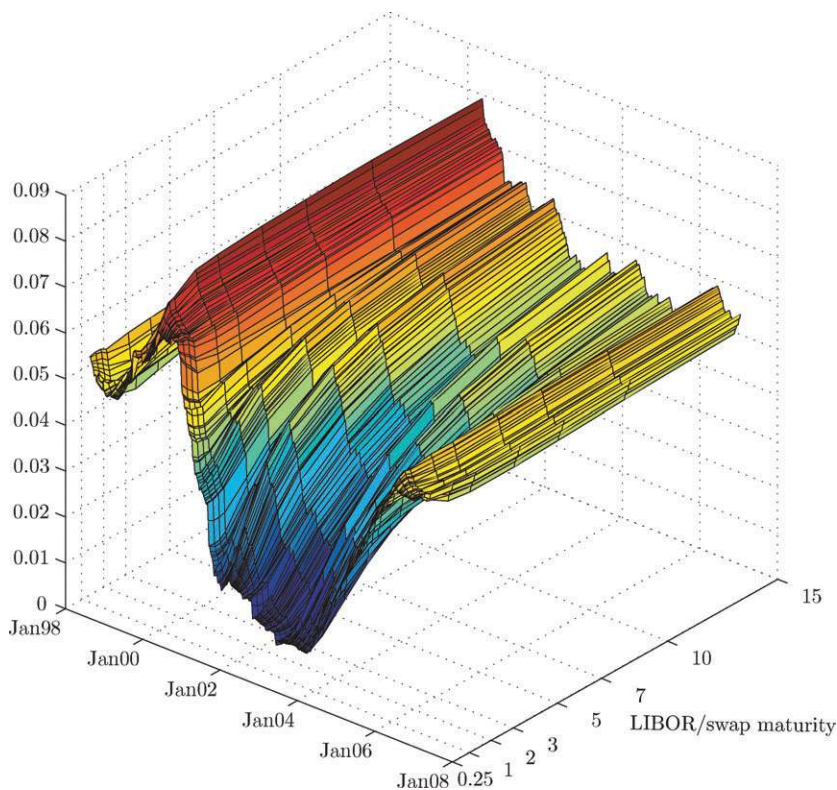
Figure 1
Time series of LIBOR and swap rates
Each time series consists of 441 weekly observations from August 21, 1998 to January 26, 2007.
Source: Bloomberg.

(we refrain from extrapolating outside the range of implied volatilities that are available and use only full skews to give equal weight to OTM and ITM caps). Furthermore, we have eliminated a few observations where there were obvious mistakes in the reported implied volatilities.

We calibrate a forward rate curve on each observation date using the following Nelson and Siegel (1987) parameterization:

$$f(t, T) = \beta_0 + \beta_1 e^{-\gamma_1(T-t)} + \beta_2(T-t)e^{-\gamma_2(T-t)}. \tag{48}$$

The parameters are recalibrated on each observation date by minimizing the mean-squared percentage differences between the observed LIBOR and swap rates on that date and those implied (48). Based on the forward rate curves (or, rather, the associated zero-coupon curves), we compute swaption and cap prices from the lognormal (or Black, 1976) pricing formulas.

For estimation, we use data from August 21, 1998 to July 8, 2005. The rest of the sample is used to evaluate the model out of sample.

Figure 2
Time series of lognormal implied ATMF swaption volatilities
Each time series consists of 441 weekly observations from August 21, 1998 to January 26, 2007.
Source: Bloomberg.

3.2 The Kalman filter

We estimate the model using the extended Kalman filter.[20] This involves writing the model in state-space form, which consists of a measurement equation and

---

[20] Duffee and Stanton (2004) compare several estimation methods in the context of estimating affine term structure models, namely Efficient Method of Moments (EMM), Simulated Maximum likelihood (SML), and
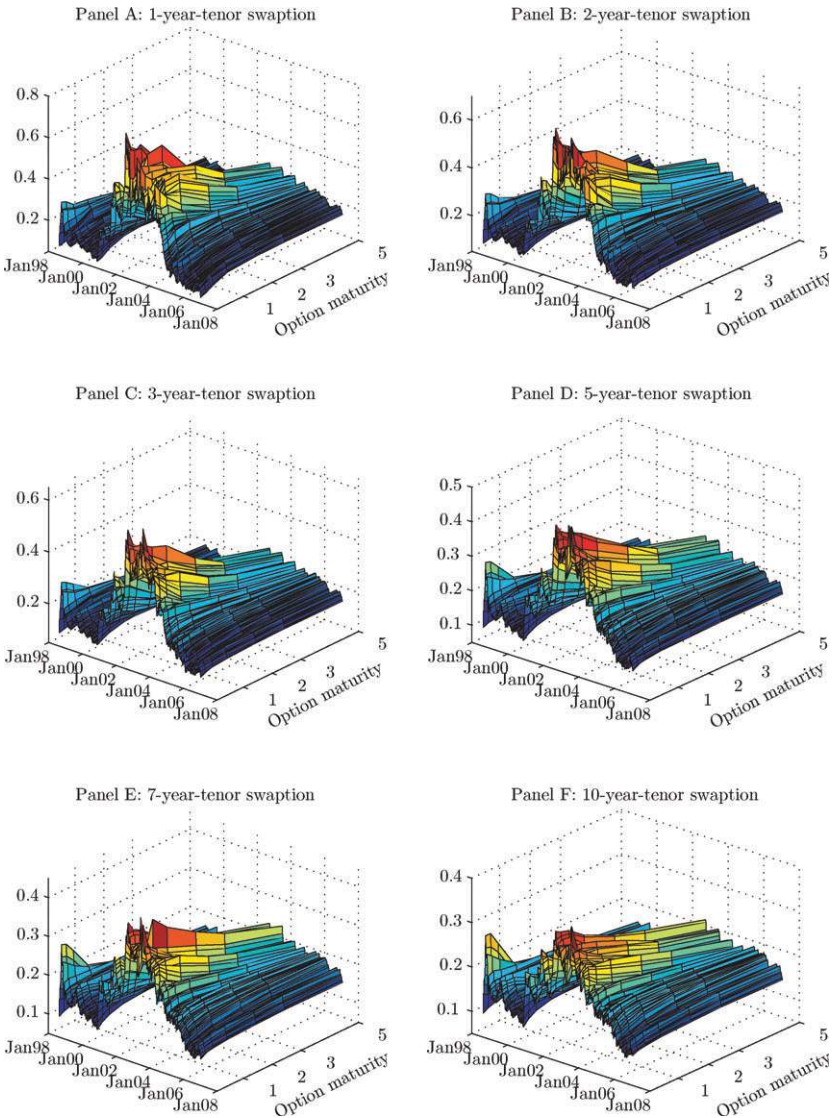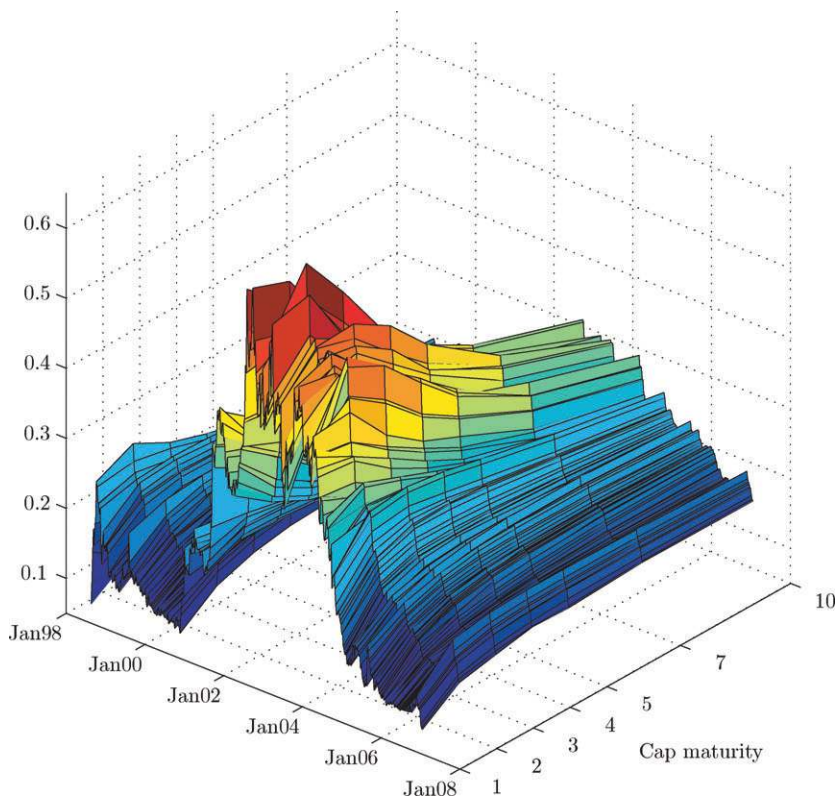
Figure 3
Time series of lognormal implied ATMF cap volatilities
Each time series consists of 441 weekly observations from August 21, 1998 to January 26, 2007.
Source: Bloomberg.

a transition equation. The measurement equation describes the relationship between observable variables and the latent state variables. It is given by

$$y_t = h(X_t) + u_t, \quad u_t \sim \text{iid. } \mathcal{N}(0, S), \tag{49}$$

where y_t is a vector consisting of observable quantities, X_t is the state vector, h is the pricing function, and u_t is a vector of iid. Gaussian measurement errors with covariance matrix S. The X_t-vector is given by

$$X_t = (x_1(t), \dots, x_N(t), \phi_{1,1}(t), \dots, \phi_{6,N}(t), v_1(t), \dots, v_N(t))', \tag{50}$$

Quasi-Maximum likelihood (QML), in conjunction with the Kalman filter. Their conclusion is that the latter procedure is preferable due to its better finite-sample properties. Computational considerations also speak in favor of the QML/Kalman filter approach, since the inclusion of derivatives in the estimation makes even this otherwise simple procedure computationally intensive. Estimating the model with more complex simulation-based EMM, SML, or MCMC procedures would be extremely time consuming, if not impossible.

Panel A: 1-year cap skew

Panel B: 2-year cap skew

Panel C: 3-year cap skew

Panel D: 4-year cap skew

Panel E: 5-year cap skew

Panel F: 7-year cap skew

Panel G: 10-year cap skew
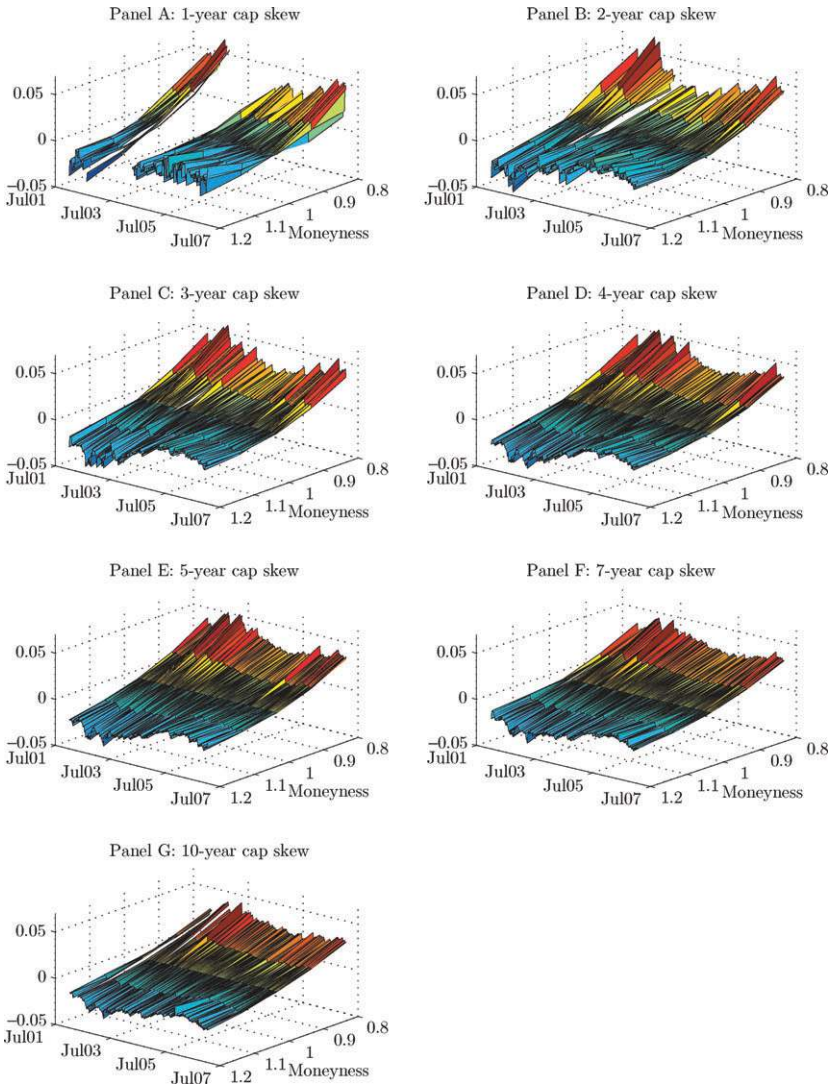
Figure 4
Time series of lognormal implied cap skews
The skews are the differences between the implied volatilities across moneyness and the implied volatilities of the corresponding ATMF caps. "Moneyness" of a given cap is defined as the ratio between its strike and the strike on the ATMF cap with the same maturity. Each time series consists of a maximum of 265 weekly observations from January 4, 2002 to January 26, 2007.
Source: Bloomberg.

while the y_t-vector consists of the LIBOR/swap term structure and the derivatives prices.

LIBOR and swap rates are nonlinearly related to $x_1(t), \ldots, x_N(t)$ and $\phi_{1,1}(t), \ldots, \phi_{6,N}(t)$ through Equation (20). The model laid out in Section 2 is

time-inhomogeneous and fits the initial yield curve by construction. For the purpose of estimation, we reduce the model to its time-homogeneous counterpart by replacing $f(0, T)$ with φ in Equation (5) and $\frac{P(0,T)}{P(0,t)}$ with $\exp\{-\varphi(T - t)\}$ in Equation (20). φ is estimated as part of the estimation procedure and can be interpreted as the infinite-maturity forward rate.[21]

Derivatives prices are nonlinearly related to $v_1(t), \ldots, v_N(t)$ through Equations (30) and (33). Since we price derivatives based on the actual forward rate curves, derivatives prices are independent of the $x(t)$ and $\phi(t)$ state variables. This has the advantage that an imperfect fit to the forward rate curve does not get reflected in derivatives prices, which in turn should provide us with a cleaner estimate of the volatility processes.[22] Since derivatives prices vary strongly across option maturities, maturities of the underlying swap rates as well as moneyness, we divide derivatives prices by their Black (1976) "vegas"— i.e., their sensitivities to variations in lognormal volatilities. With this scaling, derivatives prices have comparable magnitudes.[23]

To reduce the number of parameters in S, we make the conventional assumption that the measurement errors are cross-sectionally uncorrelated (that is, S is diagonal). Furthermore, we assume that one variance applies to all measurement errors for interest rates, and that another variance applies to all measurement errors for scaled derivatives prices.

The transition equation describes the discrete-time dynamics of the state vector implied by the continuous-time processes (44), (45), (14)–(19), $i = 1, \ldots, N$,

$$X_{t+1} = \Phi(X_t) + w_{t+1}, \quad w_{t+1} \text{ iid.}, \ E[w_{t+1}] = 0, \ \text{Var}[w_{t+1}] = Q(X_t). \quad (51)$$

Since X_t follows an affine diffusion, we have that $\Phi(X_t) = \Phi_0 + \Phi_X X_t$ and $Q(X_t) = Q_0 + \sum_{i=1}^{N} Q_{v,i} v_{t,i}$, where Φ_0, Φ_X, Q_0, and $Q_{v,i}$ are known in closed form (see, e.g., Fisher and Gilles, 1996). The disturbance vector w_{t+1} is iid. but not Gaussian.

To apply the Kalman filter, which is designed for linear Gaussian state-space models, to Equations (49) and (51), we need to linearize the h-function in Equation (49) and make the assumption that the disturbance term w_{t+1} in Equation (51) is Gaussian. With these modifications, we can apply the extended Kalman filter to Equations (49) and (51) and in the process obtain the likelihood function. For completeness the extended Kalman filter recursions are stated in Appendix C.[24] The use of a Gaussian distribution to approximate the true

---

[21] A similar approach is taken by de Jong and Santa-Clara (1999) in their estimation of HJM models.

[22] When the cap skew data is included in the estimation, the dimension of the y_t-vector varies over time. This does not present a problem, however, since the Kalman filter easily handles missing observations.

[23] This is very similar to fitting the model to lognormal implied volatilities but is much faster, since computing implied volatilities requires a numerical inversion for each swaption and cap, which would add an extra layer of complexity to the likelihood function.

[24] Classic references on the Kalman filter are Harvey (1989) and Hamilton (1994).

distribution of w_{t+1} makes this a QML estimation procedure. In Appendix C, we perform a Monte Carlo study to investigate the small-sample properties of the QML/Kalman filter approach in our setting. We find virtually no biases in the estimates of the parameters identified under Q and only small and insignificant biases in the estimates of the drift parameters in the P-dynamics.

3.3 Numerical issues

The loglikelihood function is maximized by initially using the Nelder-Mead algorithm and later switching to the gradient-based BFGS algorithm. The optimization is repeated with several different plausible initial parameter guesses to minimize the risk of not reaching the global optimum. The ODEs (31) and (32) are solved with a standard fourth-order Runge-Kutta algorithm, and the integral (34) is evaluated with the Gauss-Legendre quadrature formula, using 40 integration points and truncating the integral at 8000.[25] For the model with $N = 3$ estimated on the entire dataset up to July 8, 2005, each evaluation of the likelihood function requires calculating 60,480 swaption prices and 514,336 caplet prices,[26] underscoring the need for fast pricing routines.

4. Estimation Results

4.1 Parameter estimates

We start by estimating our model for $N = 1$, 2, and 3 on the entire dataset up to July 8, 2005. We also reestimate the model for $N = 3$ on the swaption and cap data separately to address further the relative valuation of swaptions and caps. In the following, these five models are denoted by 1SC, 2SC, 3SC, 3S, and 3C, respectively.

The five sets of parameter estimates are given in Tables 1 and 2.[27] For all the models, the estimates of $\alpha_{0,i}$, $\alpha_{1,i}$, and γ_i imply that all forward rate volatility functions are hump shaped. The need for such hump-shaped functions to match interest rate derivatives has been stressed by Amin and Morton (1994); Moraleda and Vorst (1997); Ritchken and Chuang (1999); and Mercurio and Moraleda (2000), among others, in the context of single-factor HJM models. For all the models with $N = 3$, $\sigma_{f,1}(t, T)$ affects the entire forward rate curve,

---

[25] We use 20 points on the interval 0–1000 and another 20 points on the interval 1000–8000. Increasing the number of points and/or the truncation of the integral does not change the likelihood value. In fact, truncating at 8000 is very conservative at the optimum. However, the speed with which the integrand dies out depends on the parameters and for some of the parameter vectors that are encountered during the optimization, 8000 appears an appropriate cutoff point.

[26] In the sample, there are a total of 15,120 swaptions, 43,560 caplets constituting 2520 ATMF caps, and 85,024 caplets constituting 4640 non-ATMF caps. Furthermore, the derivative (86) in Appendix C is computed numerically so we need to reprice the swaptions and caplets for small perturbations of $v_1(t)$, $v_2(t)$, and $v_3(t)$.

[27] The asymptotic covariance matrix of the estimated parameters is computed from the outerproduct of the first derivatives of the likelihood function. Theoretically, it would be more appropriate to compute the asymptotic covariance matrix from both the first and second derivatives of the likelihood function. In reality, however, the second derivatives of the likelihood function are somewhat numerically unstable.

Table 1
Parameter estimates

	N = 1 Swaptions + caps	N = 2 Swaptions + caps		N = 3 Swaptions + caps		
	i = 1	i = 1	i = 2	i = 1	i = 2	i = 3
κ_i	0.0553	0.3694	1.0364	0.5509	1.0187	0.1330
	(0.0039)	(0.0035)	(0.0142)	(0.0058)	(0.0159)	(0.0034)
σ_i	0.3325	0.8595	1.4397	1.0497	1.4274	0.5157
	(0.0091)	(0.0226)	(0.0544)	(0.0365)	(0.0432)	(0.0301)
$\alpha_{0,i}$	0.0045	0.0006	0.0004	0.0000	0.0020	−0.0097
	(0.0001)	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0003)
$\alpha_{1,i}$	0.0131	0.0071	0.0437	0.0046	0.0265	0.0323
	(0.0004)	(0.0001)	(0.0006)	(0.0001)	(0.0003)	(0.0010)
γ_i	0.3341	0.2643	1.3279	0.1777	1.1623	0.8282
	(0.0011)	(0.0008)	(0.0101)	(0.0016)	(0.0072)	(0.0028)
ρ_i	0.4615	0.2086	0.3125	0.3270	0.2268	0.1777
	(0.0320)	(0.0280)	(0.0222)	(0.0415)	(0.0161)	(0.0555)
$\kappa_{x,i}^P$	0.9767	1.0108	0.2358	0.7677	0.5650	0.8739
	(0.5280)	(0.4010)	(0.3762)	(0.6107)	(0.4014)	(0.3014)
$\kappa_{xv,i}^P$	3.4479	0.7650	1.0406	0.0988	1.7115	1.6425
	(2.4111)	(0.8154)	(0.9727)	(1.0023)	(0.8517)	(0.6079)
η_i^P	1.1964	−0.0500	0.3369	−1.1288	0.8528	1.0453
	(1.9715)	(1.5427)	(0.4361)	(2.0856)	(0.6002)	(0.3243)
κ_i^P	2.1476	1.8247	3.4793	2.3698	3.1794	1.7372
	(0.3593)	(0.4561)	(0.9697)	(0.7844)	(0.7459)	(0.1383)
θ_i^P	0.7542	1.9447	0.3890	2.1070	0.7875	0.6330
	(0.0566)	(0.2324)	(0.1047)	(0.2777)	(0.1341)	(0.2171)
φ	0.0832	0.0706			0.0680	
	(0.0003)	(0.0002)			(0.0003)	
σ_{rates}	0.0054	0.0011			0.0004	
	(0.0000)	(0.0000)			(0.0000)	
σ_{deriv}	0.0288	0.0166			0.0126	
	(0.0001)	(0.0000)			(0.0000)	
Loglikelihood	−58681.5	−41464.7			−32887.5	

Maximum-likelihood estimates with outer-product standard errors in parentheses. σ_{rates} denotes the standard deviation of interest rate measurement errors and σ_{deriv} denotes the standard deviation of swaption and cap price measurement errors. θ_i is normalized to 1. The models are estimated on weekly data from August 21, 1998 to July 8, 2005.

$\sigma_{f,2}(t, T)$ affects only the short end of the curve, and $\sigma_{f,3}(t, T)$ affects mainly the intermediate part of the curve. Panel A in Figure 5 displays the forward rate volatility functions in the case of the 3SC model.

For all the models, the first volatility state variable is more persistent than the second volatility state variable under the risk-neutral measure. Interestingly, the third volatility state variable is the most persistent for the 3SC and 3S models but the least persistent for the 3C model. This implies that shocks to the volatility state variables in the 3C model have different impacts on implied volatilities than similar shocks in the 3SC and 3C model.[28] This suggests that caps and swaptions are not priced completely consistently—an issue we return to in Section 4.4. The volatility state variables are always less persistent under P

---

[28] As discussed in Section 2.3, the impact that a $v_i(t)$-shock has on ATMF implied volatilities depends on both the $\sigma_{f,i}(t, T)$ function (and, hence, the $B_{x_i}(\tau)$ function) and the persistence of the shock. While the parameters of the $\sigma_{f,i}(t, T)$ functions are fairly similar across the 3SC, 3S, and 3C models, the persistence of $v_i(t)$-shock are not.

Table 2
Parameter estimates (cont.)

	$N = 3$ Swaptions			$N = 3$ caps		
	$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$
κ_i	0.4462	1.4196	0.2997	0.2169	0.5214	0.8340
	(0.0055)	(0.0249)	(0.0061)	(0.0236)	(0.0529)	(0.0374)
σ_i	0.9447	1.6850	0.7742	0.6586	1.0212	1.2915
	(0.0303)	(0.0530)	(0.0302)	(0.0329)	(0.0498)	(0.0484)
$\alpha_{0,i}$	−0.0000	0.0018	−0.0084	0.0000	0.0014	−0.0085
	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0002)
$\alpha_{1,i}$	0.0045	0.0191	0.0255	0.0037	0.0320	0.0272
	(0.0001)	(0.0002)	(0.0006)	(0.0001)	(0.0014)	(0.0006)
γ_i	0.1791	1.0337	0.7733	0.1605	1.4515	0.6568
	(0.0016)	(0.0062)	(0.0038)	(0.0028)	(0.0176)	(0.0065)
ρ_i	0.2720	0.2127	0.2446	0.0035	0.0011	0.6951
	(0.0759)	(0.0512)	(0.1073)	(0.0480)	(0.0128)	(0.0112)
$\kappa^P_{x,i}$	0.7410	0.4469	0.6343	0.6389	0.7539	1.1133
	(0.5811)	(0.3970)	(0.3483)	(0.4059)	(0.4392)	(0.6281)
$\kappa^P_{xv,i}$	0.0405	1.2582	1.1604	−0.1765	1.6694	1.1955
	(1.0299)	(0.7289)	(0.9510)	(0.3673)	(0.6164)	(0.8149)
η^P_i	−1.1188	1.1248	1.1100	−0.9336	0.7892	1.3072
	(2.4353)	(0.9697)	(0.3525)	(0.7350)	(0.4881)	(0.9074)
κ^P_i	2.2788	3.4535	1.6181	1.4594	3.4202	3.2223
	(0.6564)	(0.6868)	(0.3609)	(0.2192)	(0.3376)	(0.7326)
θ^P_i	2.1379	1.3648	0.8107	1.4235	0.7880	1.2602
	(0.2818)	(0.1982)	(0.1683)	(0.1706)	(0.0994)	(0.2255)
φ		0.0681			0.0668	
		(0.0002)			(0.0006)	
σ_{rates}		0.0004			0.0004	
		(0.0000)			(0.0000)	
σ_{deriv}		0.0109			0.0071	
		(0.0000)			(0.0000)	
Loglikelihood		−18947.9			−3919.2	

Maximum-likelihood estimates with outer-product standard errors in parentheses. σ_{rates} denotes the standard deviation of interest rate measurement errors and σ_{deriv} denotes the standard deviation of swaption and cap price measurement errors. θ_i is normalized to 1. The models are estimated on weekly data from August 21, 1998 to July 8, 2005.

than under Q. Panel B in Figure 5 displays the volatility state variables in the case of the 3SC model.[29]

As discussed in Section 2.1, the long-run means of the volatility state variables under Q are not identified and set to 1. All models with $N \geq 2$ have at least one volatility state variable with a long-run mean higher than 1 under P. For square-root processes, the "completely affine" risk-premium specification necessarily implies either faster mean reversion and lower long-run mean or

---

[29] In general, the stochastic state variables are highly correlated with the principal components (PCs) of the term structure and implied volatilities. In the case of the 3SC model, the correlations between changes in the three term structure state variables, $x_1(t)$, $x_2(t)$, and $x_3(t)$, and changes in the first three PCs of the LIBOR/swap term structure, often denoted "level," "slope," and "curvature" factors, are 0.941, 0.727, and 0.718, respectively. The correlations between changes in the three volatility state variables, $v_1(t)$, $v_2(t)$, and $v_3(t)$, and changes in the first three PCs of the normal implied swaption and cap volatilities are 0.911, 0.789, and 0.686, respectively. The correlations with the PCs of the lognormal implied swaption and cap volatilities are lower, which is not surprising since the volatility state variables are more directly related to the normal than the lognormal implied volatilities; see the discussion in Section 2.3. In the 3S (3C) model, the correlations between the volatility state variables and the first three PCs of the normal implied swaption (cap) volatilities are even higher.
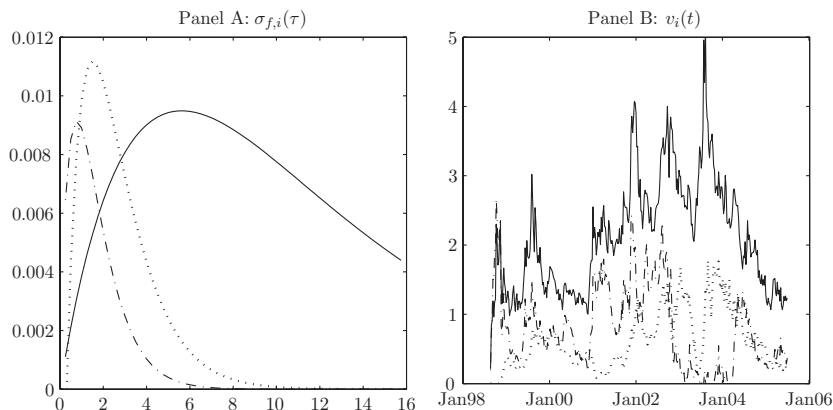
Figure 5

$\sigma_{f,i}(\tau)$ and $v_i(t)$ for the $N=3$ swaption and cap model

Panel A displays $\sigma_{f,i}(\tau)$ and Panel B displays $v_i(t)$. '———' denotes $\sigma_{f,1}(\tau)$ and $v_1(t)$, '– · –' denotes $\sigma_{f,2}(\tau)$ and $v_2(t)$, and '······' denotes $\sigma_{f,3}(\tau)$ and $v_3(t)$.

slower mean reversion and higher long-run mean under P than under Q. The combination of faster mean reversion and higher long-run mean is possible only with the "extended affine" risk-premium specification.[30]

For all the models, but the 3C model, all correlation parameters are moderately positive and statistically significant. For the 3C model, the first two correlation parameters are close to zero and insignificant while the third is positive and statistically very significant. The reason why the correlation parameters in the 3C model differ from those of the 3SC and 3S models is that shocks to the volatility state variables affect implied volatilities differently, as we have discussed above, and consequently a different set of correlation parameters is needed to match the implied cap skews and the dynamics of implied volatilities. We return to the role of the correlation parameters in Sections 4.5 and 4.6.

Finally, note that those parameters that are identified under Q are much more precisely estimated than those that are identified only under P, which is not surprising given the relatively short time series. Particularly, the drift parameters in the P-dynamics of the $x_i(t)$ state variables are very imprecisely estimated.

---

[30] In all the estimations, the boundary nonattainment condition is binding for all the volatility processes under Q but not under P. We have reestimated the models with the "completely affine" market price of risk specification, which does not impose the boundary nonattainment conditions. This yields slightly lower κ_i-estimates and somewhat higher σ_i-estimates. However, the models' pricing performances are largely unchanged. Therefore, the improvement in the models' time series fit that comes from using the "extended affine" market price of risk specification does not come at the expense of a noticeable poorer cross-sectional fit. This is consistent with results reported by Cheredito, Filipovic, and Kimmel (2007) in the context of term structure estimation without the use of derivatives.

Table 3
Model fit

			Model		
	1SC	2SC	3SC	3S	3C
			Panel A: In-sample period		
Interest rates	47.35	8.32	2.97	3.02	2.79
ATMF swaptions	10.45	6.37	4.32	3.79	12.28
ATMF caps	7.66	4.17	3.97	6.63	1.38
Non-ATMF caps	6.36	4.74	3.56	5.66	2.31
			Panel B: Out-of-sample period		
Interest rates	68.88	12.33	5.01	5.78	4.16
ATMF swaptions	11.21	8.84	4.71	3.63	14.89
ATMF caps	5.89	5.08	3.10	4.61	2.02
Non-ATMF caps	6.15	5.39	4.46	5.73	3.46

Mean of root-mean-squared pricing errors for interest rates and derivatives. For interest rates, the pricing errors are the differences between the fitted and actual interest rates. For swaptions and caps, the pricing errors are the differences between the fitted and actual prices divided by the actual prices. "1SC" denotes the $N = 1$ swaption and cap model, "2SC" denotes the $N = 2$ swaption and cap model, "3SC" denotes the $N = 3$ swaption and cap model, "3S" denotes the $N = 3$ swaption model, and "3C" denotes the $N = 3$ cap model. Interest rate pricing errors are measured in basis points while derivatives pricing errors are measured in percentages. The in-sample period is August 21, 1998 to July 8, 2005, and the out-of-sample period is July 15, 2005 to January 26, 2007.

4.2 Overall comparisons of models—in-sample and out-of-sample

For each of the estimated models, we compute the fitted LIBOR and swap rates and swaption and cap prices based on the filtered state variables. For the LIBOR and swap rates, we take the pricing errors to be the differences between the fitted and actual interest rates. For the swaptions and caps, we take the pricing errors to be the differences between the fitted and actual prices divided by the actual prices.[31] By taking the square root of the average of the squared pricing errors at each date, we construct time series of RMSEs of LIBOR/swap rates, ATMF swaptions, ATMF caps, and non-ATMF caps. Averaging these series over time produces the overall RMSEs.

We make pairwise comparisons between the models' pricing performance using the approach of Diebold and Mariano (1995). Suppose two models generate time series of root-mean-squared cap pricing errors $RMSE_{1,cap}(t)$ and $RMSE_{2,cap}(t)$. We then compute the mean of the difference $RMSE_{2,cap}(t) - RMSE_{1,cap}(t)$ and the associated t-statistics. A significantly negative mean implies that model two has a significantly better fit to caps than model one (according to the RMSE criterion).[32]

Table 3 displays the average RMSEs of LIBOR/swap rates, ATMF swaptions, ATMF caps, and non-ATMF caps for each of the five models, and Table 4 makes pairwise comparisons between the models. We report results for both the in-sample and out-of-sample periods. Consider first the in-sample period. For the

---

[31] This makes our results directly comparable to most other papers in the literature. Alternatively, we could take derivatives pricing errors to be the differences between the fitted and actual lognormal (or normal) implied volatilities.

[32] When computing the t-statistics, we use Newey and West (1987) standard errors with 12 lags to correct for heteroscedasticity and autocorrelation. The results are robust to variations in the lag length.

Table 4
Comparisons of model fit

	Model comparisons			
	2SC versus 1SC	3SC versus 2SC	3S versus 3SC	3C versus 3SC
	Panel A: In-sample period			
Interest rates	−39.03***	−5.36***	0.05	−0.18
	(−9.07)	(−9.06)	(0.79)	(−1.02)
ATMF swaptions	−4.08***	−2.05***	−0.53***	7.96***
	(−16.07)	(−13.43)	(−9.13)	(9.55)
ATMF caps	−3.50***	−0.20	2.66***	−2.59***
	(−6.70)	(−0.67)	(9.01)	(−5.13)
Non-ATMF caps	−1.62***	−1.19***	2.10***	−1.25***
	(−2.59)	(−4.42)	(8.96)	(−4.98)
	Panel B: Out-of-sample period			
Interest rates	−56.55***	−7.32***	0.77*	−0.85*
	(−15.72)	(−13.28)	(−1.66)	(−1.74)
ATMF swaptions	−2.37***	−4.13***	−1.08***	10.18***
	(−5.03)	(−10.29)	(−5.54)	(7.18)
ATMF caps	−0.81	−1.98***	1.51***	−1.08**
	(−1.30)	(−5.19)	(3.16)	(−2.52)
Non-ATMF caps	−0.76	−0.93	1.27**	−1.00***
	(1.38)	(−1.49)	(2.23)	(2.68)

Pairwise comparisons of the models' fit using the Diebold and Mariano (1995) approach. The table reports the mean differences in RMSEs with associated t-statistics in parentheses. The t-statistics are computed using Newey and West (1987) standard errors with 12 lags. "1SC" denotes the $N = 1$ swaption and cap model, "2SC" denotes the $N = 2$ swaption and cap model, "3SC" denotes the $N = 3$ swaption and cap model, "3S" denotes the $N = 3$ swaption model, and "3C" denotes the $N = 3$ cap model. Interest rate pricing errors are measured in basis points while derivatives pricing errors are measured in percentages. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The in-sample period is August 21, 1998 to July 8, 2005 and the out-of-sample period is July 15, 2005 to January 26, 2007.

1SC, 2SC, and 3SC models, the fit improves with the number of factors and the reductions in average RMSEs as N increases are generally strongly significant. These results are consistent with principal component analyses, which show that three factors are necessary to capture the variation in the term structure (see, e.g., Litterman and Scheinkman, 1991) and that additional factors unrelated to the term structure are necessary to capture the variation in ATMF swaptions (Heidari and Wu, 2003), ATMF caps (Collin-Dufresne and Goldstein, 2002a), and non-ATMF caps (Li and Zhao, 2006). The 3S model has a superior fit to swaptions, but an inferior fit to caps (which are not used for estimation) than the 3SC model. The converse holds for the 3C model, which has a superior fit to caps but an inferior fit to swaptions (which do not enter the estimation) compared with the 3SC model.[33]

The results for the out-of-sample period are similar to those of the in-sample period. The ranking of the models is the same in terms of the fit to swaptions and caps and the magnitudes of the RMSEs are similar, if only slightly larger. This is comforting as it suggests that the models do not suffer from "over-fitting."

---

[33] It appears that removing swaptions from the estimation has a bigger impact than removing caps, which to some extent has to do with the fact that there are more swaptions than caps in the sample, making the estimation procedure focus more on matching the swaption prices than cap prices when both are included in the estimation.
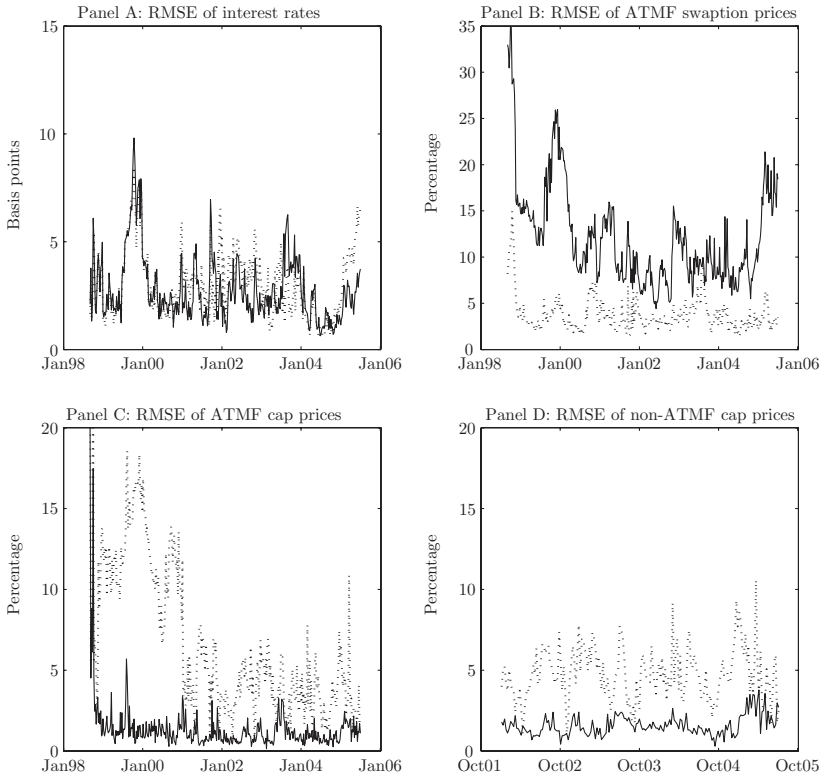
Figure 6
Time series of RMSEs for interest rates, swaptions, and caps
Panel A shows RMSEs of the basis point differences between the actual and fitted interest rates. Panel B shows RMSEs of the percentage differences between the actual and fitted ATMF swaption prices. Panel C shows RMSEs of the percentage differences between the actual and fitted ATMF cap prices. Panel D shows RMSEs of the percentage differences between the actual and fitted non-ATMF cap prices. '······' denotes the RMSEs of the $N = 3$ model fitted to term structures and swaptions. '——' denotes the RMSEs of the $N = 3$ model fitted to term structures and caps. In Panels A–C, each time series consists of 360 weekly observations from August 21, 1998, to July 8, 2005. In Panel D, each time series consists of 184 weekly observations from January 4, 2002 to July 8, 2005.

4.3 The in-sample fit to interest rates and derivatives

We now take a closer look at the fit of the models with $N = 3$. Figure 6 displays the time series of the RMSEs of LIBOR/swap rates, ATMF swaptions, ATMF caps, and non-ATMF caps for the 3S and 3C models (dotted lines and solid lines, respectively). The RMSE measure takes both variations and biases in the pricing errors into account. To see if the pricing errors for the individual interest rates and derivatives prices deviate systematically from zero, Tables 5–8 report the mean valuation errors and associated t-statistics for the LIBOR/swap rates, ATMF swaptions, ATMF caps, and non-ATMF caps, respectively, for all models with $N = 3$. We consider only the in-sample period. In this section, we consider the fit to those derivatives that enter the estimation while, in the

Table 5
Summary statistics for LIBOR and swap valuation errors

	$N = 3$ Swaptions + caps	$N = 3$ Swaptions	$N = 3$ Caps
3 months	−0.93	−1.40**	−0.60
	(−1.45)	(−2.29)	(−0.87)
6 months	−0.72*	−0.55	−0.95**
	(−1.73)	(−1.27)	(−2.40)
9 months	0.86**	1.23***	0.65*
	(2.10)	(2.98)	(1.75)
1 year	1.94***	2.26***	1.95***
	(4.22)	(5.02)	(4.56)
2 years	−1.72***	−1.95***	−1.35***
	(−3.84)	(−4.19)	(−3.02)
3 years	−0.53	−0.85**	−0.63*
	(−1.42)	(−2.25)	(−1.85)
5 years	0.59	0.59	0.26
	(1.08)	(1.12)	(0.55)
7 years	0.11	0.22	0.30
	(0.28)	(0.54)	(1.13)
10 years	0.24	0.29	0.66***
	(1.20)	(1.38)	(2.90)
15 years	−0.15	−0.14	−0.47
	(−0.20)	(−0.19)	(−0.92)

The table reports the mean pricing errors for the individual LIBOR and swap rates for each of the three $N = 3$ models. The pricing errors are defined as the differences between the fitted rates and the actual rates and are reported in basis points. T-statistics computed from Newey and West (1987) standard errors with 12 lags are in parentheses. Each statistic is computed using 360 weekly observations from August 21, 1998 to July 8, 2005. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

next section, we focus on the fit to those derivatives that are not part of the estimation (caps for the 3S model and swaptions for the 3C model).

Consider first the RMSEs. For the 3S model, the swaption RMSE (dotted line, Panel B in Figure 6) reaches about 15% in September 1998 during the LTCM crisis. Longstaff, Santa-Clara, and Schwartz (2001) and Han (2007) also report a significant increase in swaption pricing errors during this period. The swaption RMSE reaches about 10% in July 2003, when a large increase in interest rates from record low levels caused massive MBS-driven convexity hedging that also seems to have caused temporary dislocations in the derivatives market. Apart from these two episodes, the RMSE fluctuates in a range between 2% and 6%. The RMSE is comparable to that reported by Han (2007) for his preferred model with four term structure factors and three volatility factors during the sample period that overlaps with ours. Note, however, that we include a larger number of swaptions than his study. In particular, our dataset includes 1- and 3-month options and 10 year underlying swaps, which are not present in his dataset. And it is precisely these swaptions on the "edges" of the volatility surface that are the most difficult to fit.

For the 3C model, the ATMF cap RMSE (solid line, Panel C in Figure 6) also spikes in September 1998. Otherwise it mostly fluctuates between 1% and 2%. The non-ATMF cap RMSE (solid line, Panel D) also fluctuates in this range, although it breaks out of the range towards the end of the sample. The RMSE is

Table 6
Summary statistics for ATMF swaption valuation errors

Tenor	Option length						
	1mth	3mth	6mth	1yr	2yr	3yr	5yr
				$N = 3$, swaptions + caps			
1yr	−3.41***	2.52***	3.39***	3.16***	2.74***	1.92***	0.48
	(−4.37)	(3.66)	(4.75)	(6.67)	(6.46)	(4.21)	(0.67)
2yr	−0.24	0.52	2.39***	3.35***	2.95***	2.34***	1.51**
	(−0.34)	(1.09)	(5.56)	(10.07)	(9.39)	(7.02)	(2.32)
3yr	0.40	−0.13	0.63	1.86***	1.65***	1.33***	0.92
	(0.45)	(−0.19)	(1.27)	(5.13)	(5.96)	(3.82)	(1.32)
5yr	−0.73	−2.19***	−1.69***	0.02	0.56	0.41	0.14
	(−0.85)	(−3.47)	(−3.47)	(0.04)	(1.28)	(0.75)	(0.18)
7yr	1.73**	−0.39	−0.58	0.59	0.43	−0.14	−0.95
	(2.20)	(−0.76)	(−1.61)	(1.22)	(0.82)	(−0.22)	(−1.16)
10yr	3.11***	0.35	−0.41	0.13	−0.92	−2.06***	−3.36***
	(3.25)	(0.58)	(−0.86)	(0.26)	(−1.51)	(−2.93)	(−4.09)
				$N = 3$, swaptions			
1yr	−2.84***	0.80	−0.15	−1.73***	−2.00***	−1.38**	−0.93
	(−4.04)	(1.19)	(−0.20)	(−3.08)	(−3.71)	(−2.54)	(−1.19)
2yr	0.37	0.09	0.92***	0.88***	0.51	0.81**	1.09
	(0.66)	(0.30)	(3.08)	(3.10)	(1.50)	(2.13)	(1.56)
3yr	1.54**	0.34	0.42	0.96***	0.74***	1.02***	1.27*
	(2.15)	(0.71)	(1.42)	(3.33)	(2.63)	(2.68)	(1.75)
5yr	0.04	−1.70***	−1.47***	0.00	0.71*	0.99*	1.09
	(0.04)	(−2.90)	(−3.46)	(0.00)	(1.80)	(1.91)	(1.45)
7yr	1.80**	−0.41	−0.66*	0.51	0.72	0.58	0.07
	(2.14)	(−0.76)	(−1.95)	(1.23)	(1.55)	(1.02)	(0.10)
10yr	2.52**	−0.21	−0.90**	−0.17	−0.70	−1.37**	−2.40***
	(2.50)	(−0.32)	(−2.11)	(−0.40)	(−1.31)	(−2.13)	(−3.27)
				$N = 3$, caps			
1yr	−5.54***	2.32*	5.54***	8.67***	11.14***	9.11***	3.79***
	(−4.65)	(1.87)	(5.17)	(12.06)	(15.92)	(11.08)	(5.16)
2yr	5.43***	7.08***	10.17***	12.65***	12.06***	8.71***	3.87***
	(2.88)	(4.16)	(6.53)	(9.73)	(11.54)	(8.42)	(4.63)
3yr	8.64***	8.47***	9.72***	11.22***	9.07***	5.65***	2.10**
	(3.31)	(3.62)	(4.76)	(6.98)	(7.74)	(5.24)	(2.44)
5yr	1.27	0.05	0.75	2.28	1.08	−0.80	−1.38
	(0.43)	(0.02)	(0.33)	(1.19)	(0.75)	(−0.63)	(−1.32)
7yr	−4.79	−6.32**	−6.03**	−4.52**	−4.79***	−5.41***	−4.22***
	(−1.61)	(−2.37)	(−2.55)	(−2.20)	(−2.89)	(−3.75)	(−3.45)
10yr	−10.77***	−12.53***	−12.43***	−10.81***	−10.22***	−9.83***	−7.26***
	(−3.50)	(−4.55)	(−5.03)	(−4.91)	(−5.56)	(−5.99)	(−5.04)

The table reports the mean pricing errors for the individual ATMF swaptions for each of the three $N = 3$ models. The pricing errors are defined as the differences between the fitted and actual prices divided by the actual prices and are reported in percentages. T-statistics computed from Newey and West (1987) standard errors with 12 lags are in parentheses. Each statistic is computed using 360 weekly observations from August 21, 1998 to July 8, 2005. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

significantly lower than for the preferred model in Jarrow, Li, and Zhao (2007) with three term structure factors, three volatility factors, and jumps during the sample period that overlaps with ours (they report that the RMSE fluctuates around 5% during this period).

Consider next the average pricing errors in Tables 5–8. For the 3SC model, the average swaption errors range from −3.41% to 3.39% the average ATMF cap errors range from −3.08% to 0.12% and the average non-ATMF cap errors

Table 7
Summary statistics for ATMF cap valuation errors

	$N = 3$ Swaptions + caps	$N = 3$ Swaptions	$N = 3$ Caps
1 year	0.12	−2.15**	0.40
	(0.15)	(−2.01)	(0.96)
2 years	−2.14***	−6.07***	−0.01
	(−3.43)	(−7.41)	(−0.06)
3 years	−2.89***	−6.78***	0.30***
	(−4.43)	(−7.85)	(4.52)
4 years	−3.08***	−6.69***	0.33**
	(−4.38)	(−7.55)	(2.38)
5 years	−2.92***	−6.26***	0.19
	(−3.80)	(−6.81)	(1.19)
7 years	−1.95**	−4.93***	0.17
	(−2.46)	(−5.47)	(0.75)
10 years	−0.67	−3.48***	0.39
	(−0.76)	(−3.66)	(1.36)

The table reports the mean pricing errors for the individual ATMF caps for each of the three $N = 3$ models. The pricing errors are defined as the differences between the fitted and actual prices divided by the actual prices and are reported in percentages. T-statistics computed from Newey and West (1987) standard errors with 12 lags are in parentheses. Each statistic is computed using 360 weekly observations from August 21, 1998, to July 8, 2005. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

range from −4.17% to 4.15%. Quite a few of the pricing errors are statistically significant.

For the 3S model, the range of average swaption errors narrows to −2.84% to 2.52%. To put these numbers into perspective, the mean pricing errors reported by Longstaff, Santa-Clara, and Schwartz (2001) for their four-factor string market model estimated on swaptions, although for a different sample period and with their model recalibrated at every date, lie in a range from −5.37% to 5.62%.

For the 3C model, the range of average pricing errors narrows to −0.01% to 0.40% for ATMF caps and −1.51% to 1.59% for non-ATMF caps. To put these numbers into perspective, the mean pricing errors reported by Jarrow, Li, and Zhao (2007) for their preferred model estimated on cap skew data, although not for exactly the same sample period, lie in a range from −6.88% to 7.13%. Note also that, for the 3C model, far fewer of the average cap errors are statistically significant.

Finally, we briefly comment on the in-sample fit to interest rates. The RMSEs fluctuate in a range between 1 and 10 basis points, and the average errors are within a few basis points with no apparent differences between the models.

To visualize the fit, Panels A and B in Figure 7 displays the actual and fitted normal implied swaption volatility surface, on average, for the 3SC model.[34] These are clearly very similar. However, as discussed by Dai and Singleton (2002), the fitted data depend not only on the properties of a model but also on the properties of the historical data used for estimation. Therefore, comparing

---

[34] We display the swaption surface in terms of normal rather than lognormal implied volatilities since the normal implied volatilities exhibit a more pronounced hump shape that most dynamic term structure models have difficulties matching.

Table 8
Summary statistics for non-ATMF cap valuation errors

Moneyness	Cap length						
	1 year	2 years	3 years	4 years	5 years	7 years	10 years
	N = 3, swaptions + caps						
0.80	0.22	0.11	0.25	0.11	0.20	0.56***	1.31***
	(0.80)	(0.35)	(1.12)	(0.76)	(1.36)	(3.38)	(4.94)
0.90	0.20	−0.46	−0.20	−0.22*	0.03	0.61***	1.44***
	(0.51)	(−1.58)	(−1.25)	(−1.80)	(0.24)	(3.04)	(4.40)
1.00	1.55***	−1.06***	−0.73***	−0.69***	−0.28	0.30	1.50***
	(2.73)	(−3.46)	(−4.85)	(−3.93)	(−1.40)	(1.14)	(3.17)
1.10	2.77*	−2.19***	−1.76***	−1.58***	−0.96***	−0.22	1.51***
	(1.74)	(−4.36)	(−6.66)	(−6.52)	(−3.33)	(−0.60)	(2.90)
1.20	4.15	−4.17***	−3.64***	−2.83***	−2.53***	−1.20***	0.12
	(1.17)	(−6.37)	(−11.52)	(−6.41)	(−6.66)	(−3.08)	(0.17)
	N = 3, swaptions						
0.80	0.02	−1.39***	−1.56***	−1.70***	−1.57***	−1.15***	−0.40
	(0.04)	(−3.66)	(−5.77)	(−8.24)	(−7.88)	(−5.43)	(−1.37)
0.90	−0.23	−2.60***	−2.60***	−2.58***	−2.25***	−1.58***	−0.73**
	(−0.36)	(−6.36)	(−9.42)	(−12.05)	(−10.72)	(−6.29)	(−2.05)
1.00	0.38	−4.04***	−3.85***	−3.67***	−3.16***	−2.42***	−1.19**
	(0.37)	(−8.45)	(−12.35)	(−15.23)	(−12.20)	(−7.96)	(−2.41)
1.10	−0.29	−6.16***	−5.65***	−5.27***	−4.47***	−3.49***	−1.76***
	(−0.18)	(−9.75)	(−14.64)	(−18.62)	(−13.97)	(−8.43)	(−3.32)
1.20	−1.01	−9.17***	−8.36***	−7.21***	−6.66***	−5.09***	−3.69***
	(−0.34)	(−11.48)	(−17.47)	(−15.35)	(−17.51)	(−11.08)	(−5.15)
	N = 3, caps						
0.80	0.24**	0.07	0.24	−0.04	−0.17	−0.22*	0.29
	(2.02)	(0.27)	(1.00)	(−0.28)	(−1.39)	(−1.87)	(1.53)
0.90	0.05	−0.32	0.16	0.01	−0.02	−0.05	0.39***
	(0.32)	(−1.12)	(0.80)	(0.09)	(−0.31)	(−0.66)	(2.72)
1.00	0.48	−0.55***	0.31***	0.24	0.28	−0.03	0.48*
	(1.41)	(−3.22)	(5.11)	(1.31)	(1.51)	(−0.18)	(1.77)
1.10	1.29	−0.88***	0.44**	0.48	0.54	−0.04	0.62**
	(0.82)	(−3.90)	(1.97)	(1.29)	(1.36)	(−0.11)	(2.41)
1.20	1.59	−1.51***	0.31	0.76	0.31	−0.26	−0.54
	(0.46)	(−3.70)	(0.73)	(1.17)	(0.46)	(−0.72)	(−1.11)

The table reports the mean pricing errors for the individual in-the-money and out-of-the-money caps for each of the three *N* = 3 models. The pricing errors are defined as the differences between the fitted and actual prices divided by the actual prices and are reported in percentages. *T*-statistics computed from Newey and West (1987) standard errors with 12 lags are in parentheses. Each statistic is computed using a maximum of 184 weekly observations from January 4, 2002 to July 8, 2005. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

the properties of the fitted data to the actual data may in some instances yield misleading conclusions regarding the adequacy of a model. A "cleaner" way of evaluating a model is to simulate data from the model and compare the properties of the simulated data to the actual data. We, therefore, simulate (under the actual measure) 1000 samples of implied swaption volatility surfaces from the 3SC model. Each sample consists of 360 weekly observations similar to our original dataset. From these, we obtain the small-sample distribution of the average swaption volatility surface generated by the model. The mean and 95% confidence interval of this distribution are displayed in Panels C and D in Figure 7. The mean of the small-sample distribution is close to the mean of the actual data, and the mean of the actual data is certainly well within the
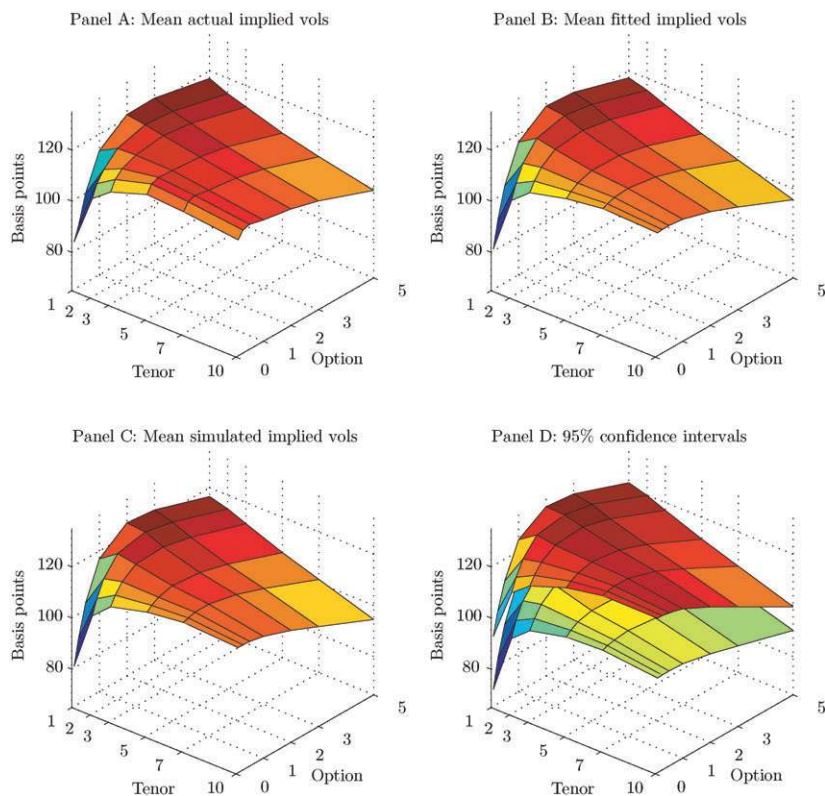
Figure 7
Means of actual, fitted, and simulated normal implied swaption volatility surfaces
Panel A shows the mean of the actual normal implied volatility surface. Panel B shows the mean of the fitted normal implied volatility surface in the case of the $N = 3$ swaption and cap model. Means are computed over 360 weekly observations from August 21, 1998 to July 8, 2005. In Panels C and D, we first simulate 1000 samples, each of length of 360, of normal implied swaption volatility surfaces. We then compute the mean volatility surface for each sample to obtain the small-sample distribution of the mean volatility surface generated by the model. Panel C shows the mean of this distribution while Panel D shows the 2.5th and 97.5th percentiles of this distribution.

95% confidence interval of the small-sample distribution.[35] This underscores the very good fit of our model.[36]

4.4 The relative valuation of caps and swaptions
We now consider the fit to those derivatives that are not part of the estimation—i.e., the fit to caps for the 3S model and the fit to swaptions for the 3C model. We

---

[35] Note that matching the mean volatility of the actual data depends crucially on the use of the "extended affine" market price of risk specification as discussed in Section 2.4.

[36] We have produced similar figures for the term structure and the normal implied ATMF cap volatility term structure. These also show the means of the small-sample distributions being close to the means of the actual data. To conserve space, we have not included these figures, but they can be found in the NBER Working Paper version of the paper.

are particularly interested in whether caps and swaptions are priced consistently with each other. In Figure 6, the out-of-sample swaption RMSE (solid line, Panel B) and out-of-sample cap RMSEs (dotted lines, Panels C and D) are larger most of the time than their in-sample counterparts. This is particularly the case in the first 2.5 years of the sample.

For the 3S model, the average cap errors in Table 7 are negative and significantly so for all caps. This means that market prices of caps have been higher on average than the prices implied by swaptions. In other words, there has been a tendency for caps to be overvalued relative to swaptions.

For the 3C model, the average swaption errors in Table 6 are significantly positive for swaptions with underlying swap maturities of 1, 2, and 3 years (except for the 1-month–into–1-year swaption) and significantly negative for swaptions with underlying swap maturities of 7 and 10 years. However, the out-of-sample results are probably most reliable for swaptions with combined swap and option maturity not exceeding 10 years, which is the maximum cap maturity in the sample. If we limit our attention to these swaptions, 25 out of 34 have positive mean pricing errors, and the mean across all 34 swaptions is 3.51%. Therefore, market prices of swaptions have generally been lower on average than the prices implied by caps. In other words, there has been a tendency for swaptions to be undervalued relative to caps consistent with the conclusions from the 3S model.

Interestingly, Longstaff, Santa-Clara, and Schwartz (2001) reach the opposite conclusion that the market has on average undervalued caps relative to swaptions, while Han (2007) finds little misvaluation on average for his preferred stochastic volatility model. These differing conclusions may to some extent be attributed to differences in models. But they may also be attributed to differences in samples, since, as we discuss next, there appear to be large variations in the relative valuation.

Figure 8 shows the average (out-of-sample) swaption valuation errors according to the 3C model (the solid line) and the average (out-of-sample) cap valuation errors according to the 3S model (the dotted line for ATMF caps and the broken line for non-ATMF caps) at each date. The figure highlights that the relative valuation between caps and swaptions fluctuates over time. According to our model, swaptions were generally overvalued relative to caps during the LTCM crisis. Subsequently, the situation reverses, and for an extended period from mid-1999 to mid-2000, swaptions appear generally undervalued relative to caps.[37] However, since then there appears to be little systematic misvaluation in the aggregate between swaptions and caps.

4.5 The role of correlation between interest rates and volatility

An important feature of our model is that it allows for nonzero correlation between innovation to forward rates and their volatilities. This is different from the

---

[37] Han (2007) also finds that for his preferred model, swaptions were undervalued relative to caps during this period, and he cites media reports that many hedge funds and proprietary traders shared this sentiment.
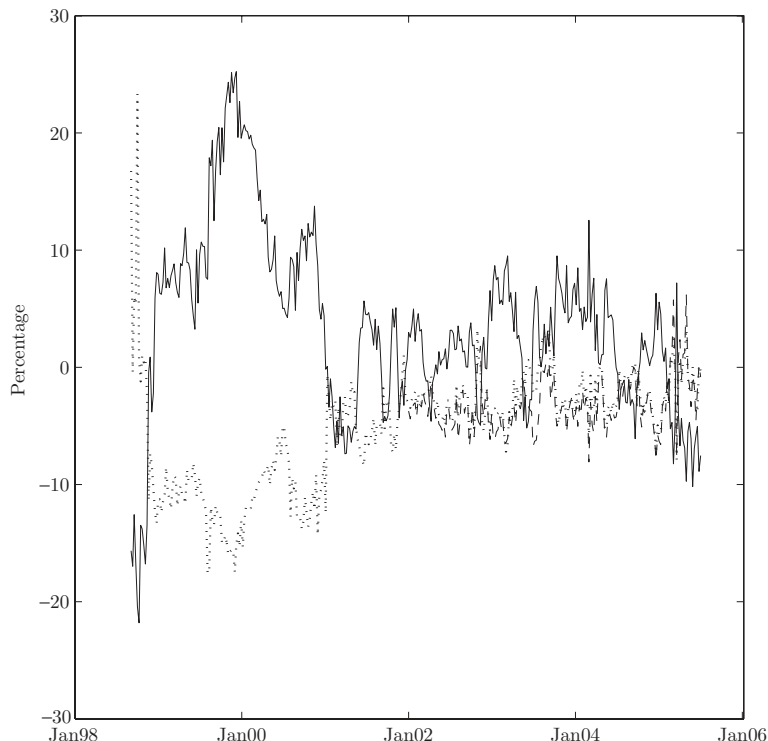
Figure 8
Time series of misvaluations of caps and swaptions
'——' denotes the average ATMF swaption valuation errors at each date according to the $N = 3$ model estimated on caps. In this case, averages are taken over swaptions with combined swap and option maturities not exceeding 10 years. '······' denotes the average ATMF cap valuation errors and '– – –' denotes the average non-ATMF cap valuation errors at each date according to the $N = 3$ model estimated on swaptions.

stochastic volatility LIBOR market models of Han (2007) and Jarrow, Li, and Zhao (2007), who impose zero correlation in order to obtain quasi-analytical option prices. Here, we discuss in more detail the role of the correlation parameters for matching the implied cap skews and the dynamics of implied volatilities.

4.5.1 Matching the implied cap skews   As discussed in Section 2.3, conditional on the volatility state variables, forward LIBOR and swap rates are approximately normally distributed in our model (under the appropriate forward measures). Suppose the correlation between innovations to the forward rate curve and volatilities were zero. In that case, the forward LIBOR rate distributions would have fatter tails than the normal distribution, and the model would predict strongly downward sloping cap skews in terms of lognormal implied volatilities, with ITM caps trading at higher lognormal implied volatilities
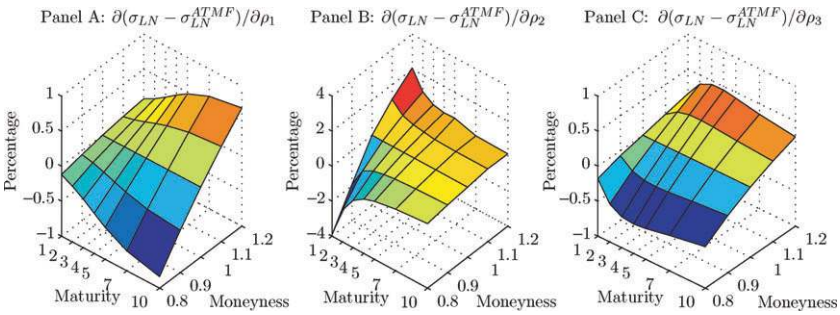
Panel A: $\partial(\sigma_{LN} - \sigma_{LN}^{ATMF})/\partial\rho_1$ Panel B: $\partial(\sigma_{LN} - \sigma_{LN}^{ATMF})/\partial\rho_2$ Panel C: $\partial(\sigma_{LN} - \sigma_{LN}^{ATMF})/\partial\rho_3$

Figure 9
The role of the ρ-parameters in matching the cap skews.
Panels A, B, and C show the derivatives of the differences between non-ATMF and ATMF lognormal implied volatilities with respect to ρ_1, ρ_2, and ρ_3, respectively. We assume that the zero-coupon curve and $v_1(t)$, $v_2(t)$, and $v_3(t)$ are initially equal to their sample averages. The responses are computed on the basis of the parameter estimates for the $N = 3$ swaption and cap model.

than OTM caps. Although this is qualitatively consistent with the data, the implied cap skews predicted by such a model will be too steep. However, the skewness of the forward LIBOR rate distributions and hence the steepness of the implied cap skews depends on the correlation parameters. To illustrate this, Figure 9 shows, for the 3SC model, the derivatives of the differences between non-ATMF and ATMF lognormal implied volatilities with respect to the correlation parameters. In all cases, increasing the correlation parameters decreases the lognormal implied volatilities of ITM caps relative to OTM caps, which decreases the steepness of the implied cap skews. It appears that ρ_1 affects mainly the implied skews of long-term caps, ρ_2 affects mainly the implied skews of short-term caps, while ρ_3 has the largest effect on implied skews of intermediate-maturity caps.

Figure 10 shows the average implied cap skews in the data (solid lines) and the average fit for the 3SC model (dashed lines), for the 3C model (dash-dotted lines), and for the 3C model reestimated with the correlation parameters restricted to zero (dotted lines). We see that the 3C model with zero correlation produces implied skews that are too steep on average. In contrast, the 3C model with nonzero correlation has an almost perfect fit to the implied skews on average. The 3SC model with nonzero correlation also has a very good fit on average although it does slightly overestimate the average steepness of the implied skews, particularly for caps of intermediate maturities.

4.5.2 Matching the dynamics of implied volatilities Figure 11, Panel A, shows the correlations between changes in lognormal implied swaption volatilities and changes in the underlying forward swap rates. For all the swaptions, the correlations are negative, more so for longer swaptions. Panel D shows the
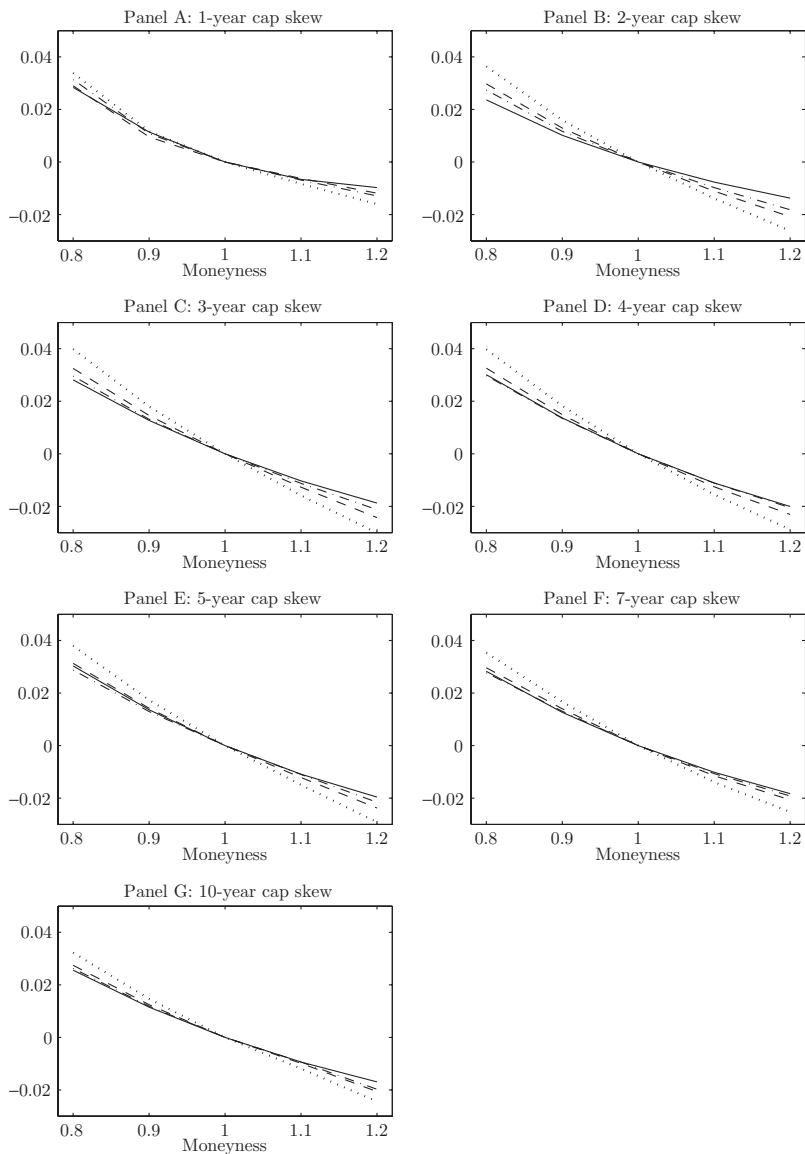
Panel A: 1-year cap skew

Panel B: 2-year cap skew

Panel C: 3-year cap skew

Panel D: 4-year cap skew

Panel E: 5-year cap skew

Panel F: 7-year cap skew

Panel G: 10-year cap skew

Figure 10
Average fit to lognormal implied cap skews
'——' denotes the average of the actual skews. '– – –' denotes the average of the fitted skews for the $N = 3$ model estimated on swaption and cap data. '– · –' denotes the average of the fitted skews for the $N = 3$ model estimated on cap data. '······' denotes the average of the fitted skews for the $N = 3$ model estimated on cap data with the correlation parameters restricted to zero. The skews are the differences between the implied volatilities across moneyness and the implied volatilities of the corresponding ATMF caps. "Moneyness" of a given cap is defined as the ratio between its strike and the strike on the ATMF cap with the same maturity. Averages are taken over a maximum of 184 weekly observations from January 4, 2002 to July 8, 2005.
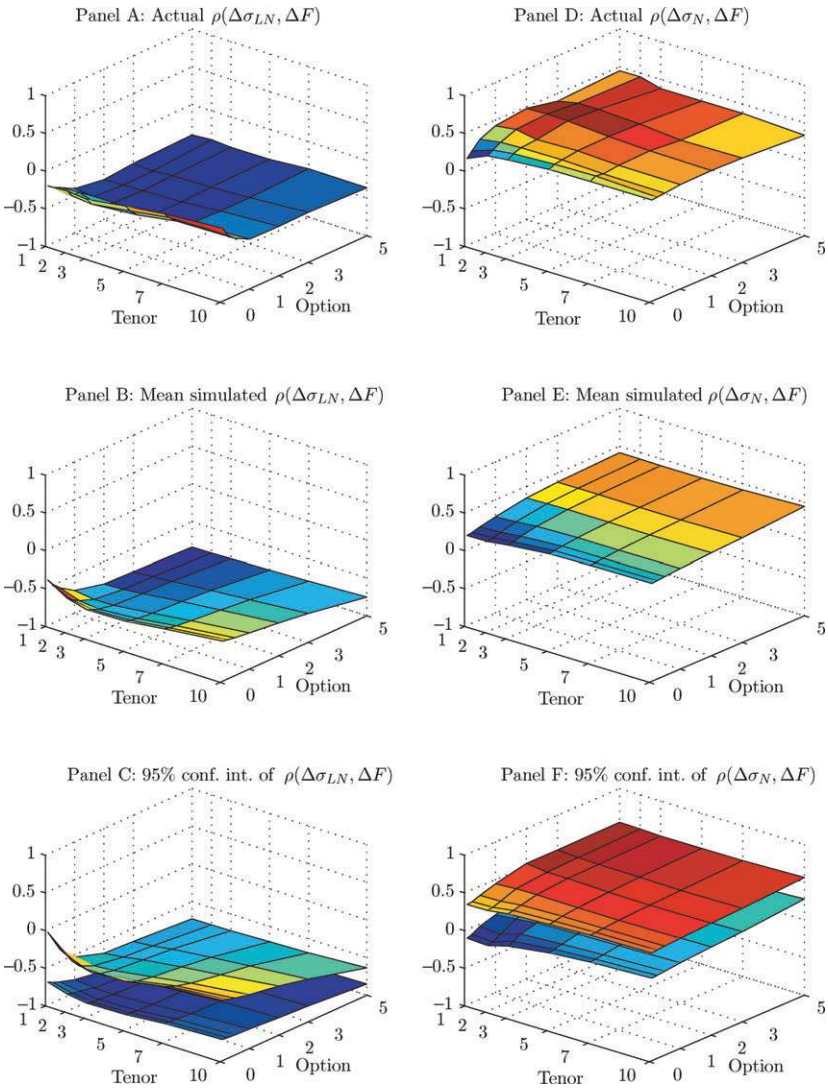Data source: Bloomberg.

Figure 11
Reproducing the implied volatility–interest rate correlations
Panel A shows the actual correlations between changes in lognormal implied swaption volatilities and changes in the underlying forward swap rates, $\rho(\Delta\sigma_{LN}, \Delta F)$. Panel D shows the actual correlations between changes in normal implied swaption volatilities and changes in the underlying forward swap rates, $\rho(\Delta\sigma_N, \Delta F)$. Each correlation is computed using 360 weekly observations from August 21, 1998 to July 8, 2005. In Panels B, C, E, and F, we first simulate 1000 samples, each of length of 360, of lognormal and normal implied volatilities and the underlying forward swap rates, in the case of the $N = 3$ swaption and model. We then compute $\rho(\Delta\sigma_{LN}, \Delta F)$ and $\rho(\Delta\sigma_N, \Delta F)$ for each sample to obtain the small-sample distributions of the correlation coefficients generated by the model. Panels B and C show the means and 2.5th and 97.5th percentiles, respectively, of the $\rho(\Delta\sigma_{LN}, \Delta F)$-distributions while Panels E and F show the means and 2.5th and 97.5th percentiles, respectively, of the $\rho(\Delta\sigma_N, \Delta F)$-distributions.

correlations using normal rather than lognormal implied swaption volatilities. In this case, all the correlations are positive.[38]

As discussed in Section 2.3, if we impose zero correlation between innovations to the forward rate curve and volatilities, the model would predict that changes in normal implied swaption volatilities were approximately uncorrelated with changes in the underlying forward rates, and that changes in lognormal implied swaption volatilities were quite strongly negatively correlated with changes in the underlying forward rates. However, with nonzero correlation, the model has more flexibility to match the actual dynamics of implied volatilities.

To see if this is the case in reality, we focus on the 3SC model and use the simulated samples discussed in Section 4.3. For each sample, we compute correlations between changes in normal and lognormal implied volatilities and changes in the underlying forward rates. This way, we obtain the small-sample distributions of the correlation coefficients generated by the model. Panels B and E in Figure 11 display the means of these distributions, while Panels C and F display the 95% confidence intervals. The model has a very good fit to the normal implied volatility correlations. The means of the small-sample distributions are generally close to the actual correlations and the actual correlations are, in any case, well within the 95% confidence bands. The model has a reasonable fit to the lognormal implied volatility correlations. The model does tend to generate too-negative correlations, but for most of the swaptions the actual correlations are within the 95% confidence bands.

4.6 Tests against nested models

The conclusion so far is that a model with three term structure factors that generate hump-shaped innovations to the forward rate curve, three additional unspanned stochastic volatility factors, and correlation between innovations to forward rates and volatility has a very good fit to the data. In this section, we investigate if the model can be simplified along certain dimensions. We reestimate the 3S and 3C models subject to the constraints $\alpha_{1,i} = 0$, $\rho_i = 1$ or $\rho_i = 0$, where $i = 1, \ldots, M$ and $M = 1, 2$, or 3.[39] The results are reported in Table 9. Panel A shows the loglikelihood values, Panel B shows the mean of root-mean-squared pricing errors for interest rates, and Panel C shows the mean of root-mean-squared pricing errors for derivatives.[40] In Table 10, we compare the restricted models with the unrestricted ones using likelihood-ratio

---

[38] These stylized facts are quite robust. For instance, computing the correlations using only the first half or the second half of the sample yields similar results. These stylized facts also hold for caps, but, to avoid an overload of figures, we concentrate on swaptions in this section.

[39] We consider both the 3S and 3C models to highlight differences between the information contained in swaptions and caps. We also reestimate the 3SC model and comment on those results when they yield additional insights.

[40] We consider only the in-sample fit, so for the 3S model derivatives pricing errors refer to ATMF swaption pricing errors, and for the 3C model derivatives pricing errors refer to ATMF and non-ATMF cap pricing errors.

Table 9
Results for restricted models

	$\alpha_{1,i} = 0$		$\rho_i = 1$		$\rho_i = 0$	
	3S	3C	3S	3C	3S	3C
			Panel A: Loglikelihood values			
Unrestricted	−18947.9	−3919.2	−18947.9	−3919.2	−18947.9	−3919.2
$i = 1$	−20523.1	−5289.4	−21656.6	−5071.4	−18948.9	−3920.3
$i = 1, 2$	−25463.0	−7856.6	−25192.4	−9116.0	−18950.3	−3922.1
$i = 1, 2, 3$	−40874.8	−16638.5	−29114.1	−15762.9	−18952.7	−5523.3
			Panel B: Interest rate RMSEs			
Unrestricted	3.02	2.79	3.02	2.79	3.02	2.79
$i = 1$	5.11	4.44	5.25	2.85	3.02	2.79
$i = 1, 2$	8.84	5.99	7.73	2.94	3.02	2.80
$i = 1, 2, 3$	13.12	6.40	14.12	8.35	3.03	2.82
			Panel C: Derivatives RMSEs			
Unrestricted	3.79	2.01	3.79	2.01	3.79	2.01
$i = 1$	3.86	2.24	4.19	2.48	3.79	2.01
$i = 1, 2$	4.82	2.64	5.12	4.36	3.80	2.03
$i = 1, 2, 3$	11.73	10.95	8.06	12.16	3.81	3.25

The table reports results from reestimating the $N = 3$ swaption model (3S) and the $N = 3$ cap model (3C) subject to the constraints $\alpha_{1,i} = 0$, $\rho_i = 1$, or $\rho_i = 0$, where $i = 1, \ldots, M$ and $M = 1, 2,$ or 3. Panel A shows the loglikelihood values. Panel B shows the mean of root-mean-squared pricing errors for interest rates with pricing errors measured as the differences between the fitted and actual interest rates. Panel C shows the mean of root-mean-squared pricing errors for derivatives with pricing errors measured as the differences between the fitted and actual prices divided by the actual prices. We consider only the in-sample fit, so for the 3S model, we consider the fit to ATMF swaptions and for the 3C model, we consider the fit to ATMF and non-ATMF caps. Interest rate pricing errors are reported in basis points while derivatives pricing errors are reported in percentages. The models are estimated on weekly data from August 21, 1998 to July 8, 2005.

tests and the Diebold and Mariano (1995) comparison of the models' pricing performance.[41]

4.6.1 Humped shaped versus exponentially declining shocks to forward rate curve The specification in Equation (4) allows the model to accommodate a wide range of shocks to the forward rate curve. In particular, it allows for hump-shaped shocks. Suppose that $\alpha_{1,1} = \cdots = \alpha_{1,M} = 0$, $M \leq N$. In this case, the first M term structure factors can generate only exponentially declining shocks to the forward rate curve. On the other hand, the dynamics of the forward rate curve simplifies considerably. It is straightforward to show that $f(t, T)$ is now given by

$$f(t, T) = f(0, T) + \sum_{i=1}^{M} \mathcal{B}_{z_i}(T - t)z_i(t) + \sum_{i=1}^{M} \mathcal{B}_{\omega_i}(T - t)\omega_i(t)$$

$$+ \sum_{i=M+1}^{N} \mathcal{B}_{x_i}(T - t)x_i(t) + \sum_{i=M+1}^{N} \sum_{j=1}^{6} \mathcal{B}_{\phi_{j,i}}(T - t)\phi_{j,i}(t), \quad (52)$$

---

[41] The likelihood-ratio test is only approximate, since the QML/Kalman filter estimation approach is not consistent in our setting. However, the Monte Carlo study in Appendix C shows the inconsistency problem to be of minor importance and we therefore believe that the likelihood-ratio test remains informative.

Table 10
Comparison between restricted and unrestricted models

	$\alpha_{1,i} = 0$		$\rho_i = 1$		$\rho_i = 0$	
	3S	3C	3S	3C	3S	3C
			Panel A: Likelihood-ratio test			
$i = 1$	1575.2***	1370.2***	2708.7***	1152.2***	2.0**	1.1
	(3150.4)	(2740.4)	(5417.4)	(2304.4)	(4.0)	(2.2)
$i = 1, 2$	6515.1***	3937.4***	6244.5***	5196.8***	3.4**	2.9*
	(13030.2)	(7874.8)	(12489.0)	(10393.6)	(6.8)	(5.8)
$i = 1, 2, 3$	21926.9***	12719.3***	10166.2***	11843.7***	4.8**	1604.1***
	(43853.8)	(25438.6)	(20332.4)	(23687.4)	(9.6)	(3208.2)
			Panel B: DM test of interest rate RMSEs			
$i = 1$	2.09***	1.65***	2.23***	0.06	0.00	0.00
	(5.14)	(5.21)	(4.42)	(1.44)	(1.01)	(0.98)
$i = 1, 2$	5.82***	3.20***	4.71***	0.15	0.00	0.01
	(10.70)	(7.39)	(10.48)	(1.23)	(1.39)	(1.35)
$i = 1, 2, 3$	10.10***	3.61***	11.10***	5.56***	0.01	0.03**
	(11.90)	(9.10)	(10.95)	(7.81)	(1.47)	(1.98)
			Panel C: DM test of derivatives RMSEs			
$i = 1$	0.07	0.23***	0.40***	0.47***	0.00	0.00
	(0.79)	(2.77)	(7.38)	(6.93)	(1.51)	(0.85)
$i = 1, 2$	1.03***	0.63***	1.33***	2.35***	0.01*	0.02
	(5.33)	(5.79)	(9.93)	(9.82)	(1.70)	(1.15)
$i = 1, 2, 3$	7.94***	8.94***	4.27***	10.15***	0.02*	1.24***
	(10.05)	(7.99)	(8.57)	(6.27)	(1.89)	(7.02)

Panel A shows, for the $N = 3$ swaption model (3S) and $N = 3$ cap model (3C), the differences in loglikelihood values between the unrestricted models and the restricted models that set $\alpha_{1,i} = 0$, $\rho_i = 1$, or $\rho_i = 0$, where $i = 1, \ldots, M$ and $M = 1, 2$, or 3. In parentheses are the likelihood-ratio test statistics. These should be compared with the critical values of a $\chi^2(M)$-distribution. Panels B and C compare the fit to interest rates and derivatives, respectively, for the restricted models with the fit for the unrestricted models using the Diebold and Mariano (1995) approach. Panel B shows the mean differences in RMSEs for interest rates, and Panel C shows the mean differences in RMSEs for derivatives (ATMF swaptions for the 3S models and ATMF and non-ATMF caps for the 3C models). In parentheses are t-statistics computed using Newey and West (1987) standard errors with 12 lags. Interest rate pricing errors are reported in basis points and derivatives pricing errors are reported in percentages. The models are estimated on weekly data from August 21, 1998 to July 8, 2005. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

where

$$\mathcal{B}_{z_i}(\tau) = \alpha_{0i} e^{-\gamma_i \tau}, \tag{53}$$

$$\mathcal{B}_{\omega_i}(\tau) = -\alpha_{0i} e^{-2\gamma_i \tau} \tag{54}$$

and $z_i(t)$ and $\omega_i(t)$ evolve according to

$$dz_i(t) = \left(\frac{\alpha_{0i}}{\gamma_i} v_i(t) - \gamma_i z_i(t) \right) dt + \sqrt{v_i(t)} dW_i^Q(t), \tag{55}$$

$$d\omega_i(t) = \left(\frac{\alpha_{0i}}{\gamma_i} v_i(t) - 2\gamma_i \omega_i(t) \right) dt, \tag{56}$$

subject to $z_i(0) = \omega_i(0) = 0$. $\mathcal{B}_{x_i}(\tau)$ and $\mathcal{B}_{\phi_{j,i}}(\tau)$ and the evolution of $x_i(t)$ and $\phi_{j,i}(t)$ are given in Proposition 1. Therefore, the dynamics of the forward rate curve are given in terms of only $M \times 3 + (N - M) \times 8$ state variables.

From Tables 9 and 10, we see that for both models the likelihood-ratio test overwhelmingly rejects the constraint $\alpha_{1,1} = \cdots = \alpha_{1,M} = 0$ even for $M = 1$.

2045

For both models, the fit to interest rates and derivatives becomes worse as M increases and the deterioration in the fit is strongly statistically significant (except for derivatives in the case of the 3S model when M increases from zero to one). The deterioration in the fit is particularly pronounced when M increases from two to three, where all volatility functions are exponentially declining. The problem with having exponentially declining shocks to the forward rate curve is that the model overestimates the volatility of short-term interest rates and hence overprices caps with short maturities and swaptions with short option maturities and short underlying swaps.[42]

4.6.2 Unspanned versus spanned volatility factors The way we set up the comparison between models in Section 4.2, a model with N term structure factors automatically allows for N unspanned stochastic volatility factors. Given that it is well established that three factors are necessary to match the dynamics of the term structure, it is perhaps not surprising that the $N = 3$ model was favored. However, it could be possible that fewer than N unspanned stochastic volatility factors are necessary to match the dynamics of interest rate derivatives. We, therefore, investigate the restriction $\rho_1 = \cdots = \rho_M = 1$, $M \leq N$. In this case, there are only $N - M$ unspanned stochastic volatility factors.

From Tables 9 and 10, we see that for both models the restriction $\rho_1 = \cdots = \rho_M = 1$ is strongly rejected by the likelihood-ratio test even for $M = 1$. This is also the case for the Diebold and Mariano (1995) test, which shows that the fit to derivatives deteriorates significantly as M increases. This confirms the studies cited above that multiple unspanned stochastic volatility factors are needed to capture fully the dynamics of interest rate derivatives.[43]

4.6.3 Nonzero versus zero correlation between interest rates and volatility In Section 4.5, we demonstrated that matching the implied cap skews and the dynamics of implied volatilities depends crucially on the correlation parameters. Here we investigate the statistical importance of nonzero correlation. In particular, we test the restriction $\rho_1 = \cdots = \rho_M = 0$, $M \leq N$. In this case, there are only $N - M$ volatility state variables that are instantaneously correlated with forward rates.

---

[42] A priori, it might seem that stochastic volatility could alleviate the need for hump-shaped forward rate volatility functions. For instance, if $v_i(t)$ is below θ_i for all i, volatility is expected to rise (under the risk-neutral measure) and, combined with exponentially declining forward rate volatility functions, the effect could be a hump-shaped implied volatility term structure (this follows from the analysis in Section 2.3), which, at least qualitatively, would be consistent with the data. However, for short-term options, the swaption surface also exhibits a hump along the swap maturity dimension; see Figure 7. Given that we have assumed independence between the term structure factors, at least one hump-shaped forward rate volatility function is needed to match the hump in this dimension.

[43] While these results show that a model with three term structure factors needs three additional unspanned stochastic volatility factors to fit interest rate derivatives, it might be the case that a model with six factors spanned by the term structure would perform even better. One approach to test this would be to compare the performance of the unrestricted $N = 3$ model with the performance of an $N = 6$ model subject to the restriction $\rho_1 = \cdots = \rho_N = 1$. Here we take a different approach. For the 3SC model, we regress innovations to each of the three unspanned factors on innovations to all LIBOR and swap rates. The R^2s equal 0.062, 0.024, and 0.037, respectively, strongly suggesting that the unspanned factors are indeed orthogonal to term structure innovations.

Consider first the likelihood-ratio test. For the 3C model, we cannot reject the restriction $\rho_1 = \cdots = \rho_M = 0$ for $M = 1$ and 2 but can strongly reject it for $M = 3$. This is not surprising since Table 2 shows two of the correlation parameter estimates being insignificant while the third is strongly significant. For the 3S model, the likelihood-ratio test does reject, although not strongly, the restriction for all M. This reflects the fact that while all the correlation parameters in Table 2 are positive, their standard deviations are larger than for the 3C model. To understand why the restriction for $M = 3$ is strongly rejected for the 3C model but only marginally rejected for the 3S model (and why the standard deviation of the correlation estimates are larger for the 3S model compared with the 3C model), recall that in principle, the correlation parameters can be identified from the variation in implied volatilities across both moneyness and time. In practice, however, the variation across moneyness provides much stronger identification than the variation across time and the former source of information is available only for the model estimated on caps.[44]

Consider next the fit to derivatives. For the 3C model, the fit to caps deteriorates significantly as M increases from 2 to 3, since the model no longer has the ability to match non-ATMF caps, as we demonstrated in Section 4.5.1. For the 3S model, the fit to swaptions is basically unchanged as M increases, which is not surprising as ATMF derivatives are virtually insensitive to the correlation parameters. However, for the 3S (and 3SC) model the fit to non-ATMF caps does become progressively worse as M increases, although this is not reported in the table since we display only in-sample results. Furthermore, if non-ATMF swaptions had been part of our sample, we would have been able, presumably, to reject the zero-correlation restrictions much more strongly for the 3S model.

5. Conclusion

We have developed a flexible stochastic volatility multifactor model of the term structure of interest rates. It features multiple unspanned stochastic volatility factors and nonzero correlation between innovations to forward rates and their volatilities. Furthermore, the model accommodates a wide range of shocks to the term structure including hump-shaped shocks. The model is highly tractable with quasianalytical prices of zero-coupon bond options and dynamics of the forward rate curve, under both the actual and risk-neutral measure, in terms of a finite-dimensional affine state vector.

We estimate the model by quasi-maximum likelihood in conjunction with the extended Kalman filter on an extensive panel dataset of LIBOR and swap rates, ATMF swaptions, ATMF caps, and non-ATMF caps (i.e., cap skews). With three term structure factors and three unspanned stochastic volatility factors, the model has a very good fit to the data. Reestimating the model on swaptions

---

[44] Consistent with these observations, for the 3SC model, the restriction $\rho_1 = \cdots = \rho_M = 0$ is strongly rejected for all M. This model has point estimates of the correlation parameters that are similar to those of the 3S model but estimated standard deviations that are comparable to those of the 3C model.

and caps separately and pricing caps and swaptions out of sample reveals that swaptions were mostly undervalued relative to caps during the first 2.5 years of the sample (at least relative to our model). However, since then swaption and cap prices appear largely consistent with each other. Testing the model against a range of alternative nested models shows that all the key features of our model are necessary to provide an adequate fit to the entire dataset.

A key result is the ability of the model to match simultaneously the implied cap skews and the dynamics of implied volatilities. This hinges on the nonzero correlation between innovations to forward rates and their volatilities.

Our model has many applications. First, the ease with which the risk-neutral dynamics of the forward rate curve can be simulated makes it useful for pricing complex interest rate derivatives by Monte Carlo simulations in which early exercise features can be handled by the Least Squares approach of Longstaff and Schwartz (2001). We believe that the model will be particularly useful for valuation of mortgage-backed securities due to its careful modeling of stochastic volatility, which is a key determinant of the value of the prepayment option.[45]

Second, with the use of the flexible "extended affine" market price of risk specification, we obtain a tractable description of the dynamics of the term structure under the actual measure, which makes the model useful in risk-management applications involving portfolios of interest rate derivatives.[46]

Third, with some adjustments, our model can be used to value derivatives on other assets. Indeed, in Trolle and Schwartz (2007), we extend the model to price commodity futures and options in a stochastic volatility HJM framework.

Appendix

A. Proofs

Proof of Proposition 1

Given Equation (4), Equation (3) becomes

$$
\mu_f(t, T) = \sum_{i=1}^{N} v_i(t) \left[\frac{\alpha_{0i}\alpha_{1i}}{\gamma_i} \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) (e^{-\gamma_i(T-t)} - e^{-2\gamma_i(T-t)}) - \frac{\alpha_{0i}\alpha_{1i}}{\gamma_i}(T-t)e^{-2\gamma_i(T-t)} \right.
$$
$$
\left. + \frac{\alpha_{1i}^2}{\gamma_i} \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) (T-t)(e^{-\gamma_i(T-t)} - e^{-2\gamma_i(T-t)}) - \frac{\alpha_{1i}^2}{\gamma_i}(T-t)^2 e^{-2\gamma_i(T-t)} \right].
$$
(57)

---

[45] Many existing MBS pricing models have difficulties matching the implied volatility skews, which in turn lead them to misprice deep-discount MBSs with significantly out-of-the-money prepayment options. The fact that our model has a good fit to the implied cap skews presumably makes it easier to match MBS prices across coupons.

[46] In a previous version of the paper, we showed that the model performs well in terms of forecasting interest rates and interest rate derivatives, beating the random walk benchmark. This depends critically on the use of the "extended affine" market price of risk specification. These results can be found in the NBER Working Paper version of the paper.

Straightforward, if slightly tedious, calculations show that

$$f(t, T) = f(0, T) + \int_0^t \mu_f(s, T)ds + \sum_{i=1}^N \int_0^t \sigma_{f,i}(s, T)\sqrt{v_i(s)}dW_i^Q(s)$$

$$= f(0, T) + \sum_{i=1}^N \mathcal{B}_{x_i}(T - t)x_i(t) + \sum_{i=1}^N \sum_{j=1}^6 \mathcal{B}_{\phi_{j,i}}(T - t)\phi_{j,i}(t), \tag{58}$$

where $\mathcal{B}_{x_i}(T - t)$ and $\mathcal{B}_{\phi_{j,i}}(T - t)$, $j = 1, \ldots, 6$ are given in the text and

$$x_i(t) = \int_0^t \sqrt{v_i(s)}e^{-\gamma_i(t-s)}dW_i^Q(s), \tag{59}$$

$$\phi_{1,i}(t) = \int_0^t \sqrt{v_i(s)}(t - s)e^{-\gamma_i(t-s)}dW_i^Q(s), \tag{60}$$

$$\phi_{2,i}(t) = \int_0^t v_i(s)e^{-\gamma_i(t-s)}ds, \tag{61}$$

$$\phi_{3,i}(t) = \int_0^t v_i(s)e^{-2\gamma_i(t-s)}ds, \tag{62}$$

$$\phi_{4,i}(t) = \int_0^t v_i(s)(t - s)e^{-\gamma_i(t-s)}ds, \tag{63}$$

$$\phi_{5,i}(t) = \int_0^t v_i(s)(t - s)e^{-2\gamma_i(t-s)}ds, \tag{64}$$

$$\phi_{6,i}(t) = \int_0^t v_i(s)(t - s)^2 e^{-2\gamma_i(t-s)}ds. \tag{65}$$

Applying Itô's Lemma to these expressions gives the dynamics stated in the text.

Proof of Proposition 2

The proof is similar to those of Duffie, Pan, and Singleton (2000) and Collin-Dufresne and Goldstein (2003). We can rewrite Equation (29) as

$$e^{-\int_0^t r_s ds}\psi(u, t, T_0, T_1) = E_t^Q\left[e^{-\int_0^{T_0} r_s ds}\psi(u, T_0, T_0, T_1)\right] \tag{66}$$

since

$$\psi(u, T_0, T_0, T_1) = e^{u\log(P(T_0, T_1))}. \tag{67}$$

Therefore, the proof consists of showing that the process

$$\eta(t) \equiv e^{-\int_0^t r_s ds}\psi(u, t, T_0, T_1) \tag{68}$$

is a martingale under Q. To this end, we conjecture that $\psi(u, t, T_0, T_1)$ is of the form (30). Applying Itô's Lemma to $\eta(t)$, and setting the drift to zero, shows that $\eta(t)$ is a martingale provided $M(\tau)$ and $N_i(\tau)$ satisfy Equations (31)–(32). Furthermore, Equation (67) holds provided that $M(0) = 0$ and $N_i(0) = 0$.

Proof of Proposition 3

Again, we follow Duffie, Pan, and Singleton (2000) and Collin-Dufresne and Goldstein (2003). The time-t price of a European put option expiring at time T_0 with strike K on a zero-coupon bond

maturing at time T_1, $\mathcal{P}(t, T_0, T_1, K)$, is given by

$$
\begin{aligned}
\mathcal{P}(t, T_0, T_1, K) &= E_t^Q\left[e^{-\int_t^{T_0} r(s)ds}(K - P(T_0, T_1))\mathbf{1}_{P(T_0,T_1)<K}\right] \\
&= KE_t^Q\left[e^{-\int_t^{T_0} r(s)ds}\mathbf{1}_{\log(P(T_0,T_1))<\log(K)}\right] \\
&\quad - E_t^Q\left[e^{-\int_t^{T_0} r(s)ds}e^{\log(P(T_0,T_1))}\mathbf{1}_{\log(P(T_0,T_1))<\log(K)}\right] \\
&= KG_{0,1}(\log(K)) - G_{1,1}(\log(K)),
\end{aligned}
\tag{69}
$$

where

$$
G_{a,b}(y) = E_t^Q\left[e^{-\int_t^{T_0} r(s)ds}e^{a\log(P(T_0,T_1))}\mathbf{1}_{b\log(P(T_0,T_1))<y}\right].
\tag{70}
$$

To evaluate $G_{a,b}(y)$, note that its Fourier transform is given by

$$
\begin{aligned}
\mathcal{G}_{a,b}(y) &= \int_{\mathbb{R}} e^{\mathrm{i}uy}\,dG_{a,b}(y) \\
&= E_t^Q\left[e^{-\int_t^{T_0} r_s ds}e^{(a+\mathrm{i}ub)\log(P(T_0,T_1))}\right] \\
&= \psi(a + \mathrm{i}ub, t, T_0, T_1),
\end{aligned}
\tag{71}
$$

where $\mathrm{i} = \sqrt{-1}$. Applying the Fourier inversion theorem, we have

$$
G_{a,b}(y) = \frac{\psi(a, t, T_0, T_1)}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}[\psi(a + \mathrm{i}ub, t, T_0, T_1)e^{-\mathrm{i}uy}]}{u}\,du.
\tag{72}
$$

B. Pricing of Swaps, Caps, and Swaptions

LIBOR and swap rates
The time-t LIBOR rate for the period t to T is given by

$$
L(t, T) = \frac{1 - P(t, T)}{(T - t)P(t, T)}.
\tag{73}
$$

Similarly, the time-t forward LIBOR rate for the period T_1 to T_2 is given by

$$
L(t, T_1, T_2) = \frac{P(t, T_1) - P(t, T_2)}{(T_2 - T_1)P(t, T_2)}.
\tag{74}
$$

In the following, consider a period length ν and a set of dates $T_j = t + \nu j$, $j = 1, \ldots, n$. The time-t swap rate for the period t to T_n and fixed-leg payments at $T_1, \ldots, T_n$[47] is given by

$$
S(t, T_n) = \frac{1 - P(t, T_n)}{\nu \sum_{j=1}^n P(t, T_j)}.
\tag{75}
$$

---

[47] Market convention for USD swaps is semiannual fixed-leg payments (but quarterly floating-leg payments). Therefore $\nu = 1/2$ in swap rate calculations. To ease notation, we assume a constant period length between payment dates. In reality, USD LIBOR derivatives are quoted on an Actual/360 basis, and in all computations in the paper, we take into account the slightly varying period length between reset dates.

Similarly, the time-t forward swap rate for the period T_m to T_n and fixed-leg payments at $T_{m+1}, \ldots, T_n$ is given by

$$S(t, T_m, T_n) = \frac{P(t, T_m) - P(t, T_n)}{\nu \sum_{j=m+1}^{n} P(t, T_j)}. \tag{76}$$

Cap prices

A cap with payments at $T_2, \ldots, T_n$ consists of $n - 1$ caplets expiring at these dates.[48] A caplet expiring at T_j with strike K is a call option on the LIBOR rate $L(t, T_j - \nu, T_j)$ with the payoff

$$\chi = \nu(L(T_j - \nu, T_j) - K)^+, \tag{77}$$

where K is the strike. This payoff is fixed at time $T_j - \nu$. Easy manipulations show that the price of the caplet at time $T_j - \nu$ equals $(1 + \nu K)$ European put options, expiring at time $T_j - \nu$, with strike $1/(1 + \nu K)$ written on a zero-coupon bond maturing at time T_j. The caplet price at time $t < T_j - \nu$ is therefore given as

$$\text{Cpl}(t, T_j - \nu, T_j, K) = (1 + \nu K)\mathcal{P}\left(t, T_j - \nu, T_j, \frac{1}{1 + \nu K}\right). \tag{78}$$

The time-t price of a cap with strike $\overline{K}$ and payments at $T_2, \ldots, T_n$ is the sum of the prices of the $n - 1$ underlying caplets, i.e.,

$$\text{Cap}(t, T_n, \overline{K}) = \sum_{j=2}^{n} \text{Cpl}\left(t, T_j - \nu, T_j, \overline{K}\right). \tag{79}$$

When the strike $\overline{K}$ equals the time-t forward swap rate for the period T_1 to T_n with fixed-leg payments at $T_2, \ldots, T_n$, the price of the cap equals the price of the corresponding floor and the cap and the floor are ATMF.

Swaption prices by the stochastic duration approach

The time-t value of a forward payer swap for the period T_m to T_n with fixed-leg payments at $T_{m+1}, \ldots, T_n$ and fixed rate K is given by

$$V(t, T_m, T_n) = P(t, T_m) - P(t, T_n) - K\nu \sum_{j=m+1}^{n} P(t, T_j). \tag{80}$$

A payer swaption is an option to enter into a payer swap at a given fixed rate. A $(T_m - t)$–into–$(T_n - T_m)$ payer swaption (i.e., an option expiring at T_m on a payer swap for the period T_m to T_n), with strike K has a payoff at T_m of

$$\chi = V(T_m, T_m, T_n)^+ = \left(1 - P(T_m, T_n) - K\nu \sum_{j=m+1}^{n} P(T_m, T_j)\right)^+. \tag{81}$$

---

[48] Market convention for USD caps is quarterly payments. Therefore $\nu = 1/4$ in cap and caplet price calculations. Furthermore, market convention is to exclude the first caplet expiring at T_1 from the cap since its cash flow is known already at time t.

Hence, a payer swaption can be viewed as a put option with strike 1 on a coupon bond with coupon rate K. In the following, we let $P^c(t)$ denote the time-t value of this coupon bond. That is,

$$P^c(t) = \sum_{j=m+1}^{n} Y(T_j)P(t, T_j), \tag{82}$$

where $Y(T_j) = K\nu$ for $j = m + 1, \ldots, n - 1$ and $Y(T_n) = 1 + K\nu$.

The stochastic duration $D(t)$ of a coupon bond is the maturity of the zero-coupon bond, that has the same relative volatility as the coupon bond. In our model, the stochastic duration of $P^c(t)$ is therefore given implicitly as the solution to

$$\sum_{i=1}^{N} v_i(t)B_{x_i}(D(t))^2 = \sum_{i=1}^{N} v_i(t)\left(\sum_{j=m+1}^{n} w_j B_{x_i}(T_j - t)\right)^2, \tag{83}$$

where $w_j = \frac{Y(T_j)P(t, T_j)}{\sum_{j=m+1}^{n} Y(T_j)P(t, T_j)}$. $D(t)$ must be found numerically, but it exists and is unique if $B_{x_i}(\tau)$ is uniformly decreasing. See Munk (1999).[49]

Wei (1997) and Munk (1999) suggest approximating an option on a coupon bond with an (scaled) option on a zero-coupon bond with maturity equal to the stochastic duration of the coupon bond. Let Swpn(t, T_m, T_n, K) denote the time-t price of a $(T_m - t)$–into–$(T_n - T_m)$ payer swaption with strike K. According to Wei (1997) and Munk (1999) this swaption price is approximately given by

$$\text{Swpn}(t, T_m, T_n, K) = \zeta \mathcal{P}(t, T_m, t + D(t), \zeta^{-1}), \tag{84}$$

where $\zeta = \frac{P^c(t)}{P(t, t+D(t))}$ is a scaling factor.

C. Estimation Details

The extended Kalman filter

Let $\hat{X}_t = E_t[X_t]$ and $\hat{X}_{t|t-1} = E_{t-1}[X_t]$ denote expectations of X_t (respectively including and excluding y_t), and let P_t and $P_{t|t-1}$ denote the corresponding estimation error covariance matrices. Linearizing the h-function in Equation (49) around $\hat{X}_{t|t-1}$, we obtain

$$y_t = (h(\hat{X}_{t|t-1}) - H'_t \hat{X}_{t|t-1}) + H'_t X_t + u_t, \quad u_t \sim \text{iid. } \mathcal{N}(0, S), \tag{85}$$

where

$$H'_t = \left.\frac{\delta h(X_t)}{\delta X'_t}\right|_{X_t = \hat{X}_{t|t-1}}. \tag{86}$$

Assuming w_t in Equation (51) is Gaussian, we obtain

$$X_t = \Phi_0 + \Phi_X X_{t-1} + w_t, \quad w_t \sim \text{iid. } \mathcal{N}(0, Q_t). \tag{87}$$

The Kalman filter applied to Equations (85) and (87) is given by the following recursions:

$$\hat{X}_{t|t-1} = \Phi_0 + \Phi_X \hat{X}_{t-1}, \tag{88}$$

$$P_{t|t-1} = \Phi_X P_{t-1} \Phi'_X + Q_t \tag{89}$$

---

[49] At the parameter estimates reported in Table 1 and 2, some $B_{x_i}(\tau)$ exhibit a tiny hump for very short maturities. However, over the relevant range of maturities all $B_{x_i}(\tau)$ are indeed uniformly decreasing.

and

$$\hat{X}_t = \hat{X}_{t|t-1} + P_{t|t-1} H_t' F_t^{-1} \epsilon_t, \qquad (90)$$

$$P_t = P_{t|t-1} - P_{t|t-1} H_t' F_t^{-1} H_t P_{t|t-1}, \qquad (91)$$

where

$$\epsilon_t = y_t - h(\hat{X}_{t|t-1}), \qquad (92)$$

$$F_t = H_t P_{t|t-1} H_t' + S. \qquad (93)$$

The loglikelihood function is constructed from Equations (92) and (93) as

$$\log L = -\frac{1}{2}\log(2\pi)\sum_{i=1}^{T} N_t - \frac{1}{2}\sum_{i=1}^{T}\log|F_t| - \frac{1}{2}\sum_{i=1}^{T}\epsilon_t' F_t^{-1}\epsilon_t, \qquad (94)$$

where T is the number of observation dates and N_t is the dimension of ϵ_t. We follow standard practice in the literature and initialize the Kalman filter at the unconditional mean and covariance matrix of X_t.

The small-sample properties of the QML/Kalman filter approach

While QML estimation has been shown to be consistent in many settings, it is in fact not consistent when used in conjunction with the extended Kalman filter.[50] More importantly, even if it were consistent, the estimator might exhibit small-sample biases. Here we perform a Monte Carlo analysis to assess the small-sample properties of the QML/Kalman filter approach in our setting. It is based on 200 simulated data sets, each of which has the same length as the original dataset—i.e., 360 observations at a weekly frequency. We perform only the analysis for the $N = 1$ model since it is computationally burdensome to reestimate the model a large number of times.

To generate each dataset, we first simulate the state vector using a simple Euler discretization of the process with 50 intermediate steps per week. The parameters are from Table 1 and the simulation is initiated at the unconditional mean of the state vector.[51] Second, we compute the same LIBOR and swap rates, swaption and cap prices as in the original dataset.[52] Third, we add normally distributed and serially and cross-sectionally uncorrelated measurement errors to the observations.[53]

Table 11 displays the results. It shows the true parameters, the asymptotic "outer-product" standard errors, as well as the mean, median, and standard deviation of the estimates from the Monte Carlo simulations. We see that there is virtually no bias in estimates of the parameters identified under Q. Estimates of the parameters identified only under P do display some biases, although these are small relative to the standard deviations of the estimates, rendering the biases insignificant. These findings are consistent with those of other studies that investigate the small-sample properties of the QML/Kalman filter approach in the context of estimating stochastic volatility term structure models without derivatives—see, e.g., Lund (1997); Duan and Simonato (1999); and Duffee and Stanton (2004). Note that for the parameters in the Q dynamics, the asymptotic standard errors are

---

[50] This is due to the linearization of the h function, and the fact that the conditional covariance matrix Q in the recursions depends on the Kalman filter estimate $\widehat{v}_t$ rather than the true, but unobservable, v_t; see Duan and Simonato (1999) and Lund (1997) for more details on this issue.

[51] An initial 200 observations are generated and then discarded to avoid dependence on the initial values.

[52] To be consistent with the original dataset, we compute only non-ATMF cap prices for the last 184 observations.

[53] We use the estimate for the standard deviations of the measurement errors reported in Table 1, i.e., 54 bp for interest rates and $0.0288 \times$ "vega" for derivatives prices.

Table 11
Small-sample properties of the QML/Kalman filter approach

	True parameter	MC mean	MC median	"o-p" standard error	MC standard deviation
κ_1	0.0553	0.0587	0.0588	0.0039	0.0076
σ_1	0.3325	0.3418	0.3429	0.0091	0.0220
$\alpha_{0,1}$	0.0045	0.0045	0.0045	0.0001	0.0001
$\alpha_{1,1}$	0.0131	0.0130	0.0130	0.0004	0.0003
γ_1	0.3341	0.3340	0.3338	0.0011	0.0021
ρ_1	0.4615	0.4633	0.4656	0.0320	0.0315
$\kappa_{x,1}^P$	0.9767	1.0142	1.0318	0.5280	0.4412
$\kappa_{xv,1}^P$	3.4479	2.5045	2.7230	2.4111	1.1938
η_1^P	1.1964	1.4046	1.3823	1.9715	1.2136
κ_1^P	2.1476	2.2534	2.1953	0.3593	0.2663
θ_1^P	0.7542	0.8091	0.8054	0.0566	0.0665
φ	0.0832	0.0831	0.0831	0.0003	0.0005
σ_{rates}	0.0054	0.0054	0.0054	0.0000	0.0001
σ_{deriv}	0.0288	0.0288	0.0288	0.0001	0.0002

We simulate 200 data sets from the $N = 1$ swaption and cap model with parameters equal to those given in Table 1. Then, we reestimate the model on each of the simulated data sets using the QML/Kalman filter approach. The table shows the true parameter values followed by the mean and median estimates from the simulated data sets. It then shows the asymptotic "outer-product" standard errors reported in Table 1 and the standard deviation of the estimates across the simulated data sets.

fairly close to the finite-sample standard errors, while for the drift parameters in the P dynamics, the asymptotic standard errors are typically larger than the finite-sample standard errors.

Although we have reported only results for the $N = 1$ model, we expect similar results to hold for the $N = 2$ and $N = 3$ models. While these models have more parameters, they also have smaller measurement errors, and the performance of the extended Kalman filter typically improves with the "signal-to-noise" ratio—see, e.g., Pichard (1991).

References
Almeida, C., J. Graveline, and S. Joslin. 2006. Do Interest Rate Options Contain Information about Excess Returns. Working Paper, Stanford Graduate School of Business.

Amin, K., and A. Morton. 1994. Implied Volatility Functions in Arbitrage-Free Term Structure Models. *Journal of Financial Economics* 35:141–80.

Andersen, L., and R. Brotherton-Ratcliffe. 2005. Extended Libor Market Models with Stochastic Volatility. *Journal of Computational Finance* 9:1–40.

Andersen, T. G., and L. Benzoni. 2005. Can Bonds Hedge Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models. Working Paper, Kellogg School of Management, Northwestern University.

Andersen, T. G., and J. Lund. 1997. Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate. *Journal of Econometrics* 77:343–77.

Ball, C. A., and W. N. Torous. 1999. The Stochastic Volatility of Short-Term Interest Rates: Some International Evidence. *Journal of Finance* 54:2339–59.

Bhar, R., and C. Chiarella. 1997. Transformation of Heath-Jarrow-Morton Models to Markovian Systems. *European Journal of Finance* 3:1–26.

Bikbov, R., and M. Chernov. 2004. Term Structure and Volatility: Lessons from the Eurodollar Markets. Working Paper, Columbia University.

Black, F. 1976. The Pricing of Commodity Contracts. *Journal of Financial Economics* 3:167–79.

Brace, A., D. Gatarek, and M. Musiela. 1997. The Market Model of Interest Rate Dynamics. *Mathematical Finance* 7:127–55.

Casassus, J., P. Collin-Dufresne, and R. Goldstein. 2005. Unspanned Stochastic Volatility and Fixed Income Derivatives Pricing. *Journal of Banking and Finance* 29:2723–49.

Chan, K., G. Karolyi, F. Longstaff, and A. Sanders. 1992. An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance* 47:1209–27.

Chen, R.-R., and L. Scott. 2001. Stochastic Volatility and Jumps in Interest Rates: An Empirical Analysis. Working Paper, Rutgers University.

Cheredito, P., D. Filipovic, and R. Kimmel. 2007. Market Price of Risk Specifications for Affine Models: Theory and Evidence. *Journal of Financial Economics* 83:123–70.

Chiarella, C., and O. K. Kwon. 2003. Finite Dimensional Affine Realisations of HJM Models in Terms of Forward Rates and Yields. *Review of Derivatives Research* 5:129–55.

Collin-Dufresne, P., and R. Goldstein. 2002a. Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. *Journal of Finance* 57:1685–30.

Collin-Dufresne, P., and R. Goldstein. 2002b. Pricing Swaption within an Affine Framework. *Journal of Derivatives* 10:1–18.

Collin-Dufresne, P., and R. Goldstein. 2003. Generalizing the Affine Framework to HJM and Random Field Models. Working Paper, U.C. Berkeley.

Collin-Dufresne, P., R. Goldstein, and C. Jones. 2003. Identification and Estimation of 'Maximal' Affine Term Structure Models: An Application to Stochastic Volatility. Working Paper, U.C. Berkeley.

Dai, Q., and K. Singleton. 2000. Specification Analysis of Affine Term Structure Models. *Journal of Finance* 55:1943–78.

Dai, Q., and K. Singleton. 2002. Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure. *Journal of Financial Economics* 63:415–42.

Dai, Q., and K. Singleton. 2003. Term Structure Dynamics in Theory and Reality. *Review of Financial Studies* 16:631–78.

de Jong, F., and P. Santa-Clara. 1999. The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables. *Journal of Financial and Quantitative Analysis* 34, 131–57.

Diebold, F., and R. Mariano. 1995. Comparing Predctive Accuracy. *Journal of Business and Economic Statistics* 13, 253–63.

Duan, J.-C., and J.-G. Simonato. 1999. Estimating and Testing Eksponential-Affine Term Structure Models by Kalman Filter. *Review of Quantitative Finance and Accounting* 13, 111–35.

Duffee, G. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* 57:405–43.

Duffee, G., and R. Stanton. 2004. Estimation of Dynamic Term Structure Models. Working Paper, U.C. Berkeley.

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

Duffie, D., J. Pan, and K. Singleton. 2000. Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* 68:1343–76.

Fisher, M., and C. Gilles. 1996. Estimating Exponential-Affine Models of the Term Structure. Working Paper, Federal Reserve Board.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, N. J.: Princeton University Press.

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

Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Heath, D., R. Jarrow, and A. Morton. 1992. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* 60:77–05.

Heidari, M., and L. Wu. 2003. Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates? *Journal of Fixed Income* 13:75–86.

Heston, S. 1993. A Closed Form Solution for Options with Stochastic Volatility. *Review of Financial Studies* 6:327–43.

Ho, T., and S. Lee. 1986. Term Structure Movements and Pricing Interest Rate Contingent Claims. *Journal of Finance* 41:1011–29.

Hull, J., and A. White. 1990. Pricing Interest-Rate-Derivative Securities. *Review of Financial Studies* 3:573–92.

Inui, K., and M. Kijima. 1998. A Markovian Framework in Multi-Factor Heath-Jarrow-Morton Models. *Journal of Financial and Quantitative Analysis* 33, 423–40.

Jamshidian, F. 1997. Libor and Swap Market Models and Measures. *Finance and Stochastics* 1:293–330.

Jarrow, R., H. Li, and F. Zhao. 2007. Interest Rate Caps 'Smile' Too! But Can the LIBOR Market Models Capture the Smile?. *Journal of Finance* 62:345–82.

Li, H., and F. Zhao. 2006. Unspanned Stochastic Volatlity: Evidence from Hedging Interest Rate Drivatives. *Journal of Finance* 61:341–78.

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

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

Longstaff, F., and E. Schwartz. 1992. Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model. *Journal of Finance* 47:1259–82.

Longstaff, F., and E. Schwartz. 2001. Valuing American Options by Simulation: A Simple Least-Square Approach. *Review of Financial Studies* 14:113–47.

Lund, J. 1997. Econometric Analysis of Continuous-Time Arbitrage-Free Models of the Term Structure of Interest Rates. Working Paper, Aarhus School of Business.

Mercurio, F., and J. Moraleda. 2000. An Analytically Tractable Interest Rate Model with Humped Volatility. *European Journal of Operational Research* 120:205–14.

Miltersen, K., K. Sandmann, and D. Sondermann. 1997. Closed Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates. *Journal of Finance* 52:409–30.

Moraleda, J., and T. Vorst. 1997. Pricing American Interest Rate Claims with Humped Volatility Models. *Journal of Banking and Finance* 21:1131–57.

Munk, C. 1999. Stochastic Duration and Fast Coupon Bond Option Pricing in Multi-Factor Models. *Review of Derivatives Research* 3:157–81.

Nelson, C., and A. Siegel. 1987. Parsimonious Modeling of Yield Curves. *Journal of Business* 60:473–89.

Newey, W., and K. West. 1987. A Simple, Positive Semi-Definit, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–08.

Piazzesi, M. 2005. Bond Yields and the Federal Reserve. *Journal of Political Economy* 113:311–44.

Pichard, J. 1991. Efficiency of the Extended Kalman Filter for Nonlinear Systems with Small Noise. *SIAM Journal on Applied Mathematics* 51:843–85.

Ritchken, P., and I. Chuang. 1999. Interest Rate Option Pricing with Volatility Humps. *Review of Derivatives Research* 3:237–62.

Ritchken, P., and L. Sankarasubramaniam. 1995. Volatility Structures of Forward Rates and the Dynamics of the Term Structure. *Mathematical Finance* 5:55–72.

Schrager, D., and A. Pelsser. 2006. Pricing Swaption in Affine Term Structure Models. *Mathematical Finance* 16:673–94.

Singleton, K., and L. Umantsev. 2002. Pricing Coupon-Bond Options and Swaptions in Affine Term Structure Models. *Mathematical Finance* 12:427–46.

Trolle, A. B., and E. S. Schwartz. 2007. Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives. Working Paper, UCLA and NBER.

Umantsev, L. 2001. Econometric Analysis of European Libor-Based Options within Affine Term Structure Models. Ph.D. Dissertation, Stanford University.

Wei, J. 1997. A Simple Approach to Bond Option Pricing. *Journal of Futures Markets* 17:131–60.

Wu, L., and F. Zhang. 2005. Fast Swaption Pricing Under a Market Model with Stochastic Volatility. Working Paper, Hong Kong University of Science and Technology.