

NBER WORKING PAPER SERIES

A GENERAL STOCHASTIC VOLATILITY MODEL FOR THE
PRICING AND FORECASTING OF INTEREST RATE DERIVATIVES

Anders B. Trolle
Eduardo S. Schwartz

Working Paper 12337
<http://www.nber.org/papers/w12337>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2006

We thank Francis Longstaff and seminar participants at UCLA Anderson School of Management for comments. Eduardo Schwartz: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095-1481. E-mail: eduardo.schwartz@anderson.ucla.edu. Anders Trolle: Copenhagen Business School, Solbjerg Plads 3, A5, DK-2000 Frederiksberg, Denmark. E-mail: abt.fi@cbs.dk. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2006 by Anders B. Trolle and Eduardo S. Schwartz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Anders B. Trolle and Eduardo S. Schwartz

NBER Working Paper No. 12337

June 2006

JEL No. E43, G13

ABSTRACT

We develop a tractable and flexible stochastic volatility multi-factor model of the term structure of interest rates. It features correlations between innovations to forward rates and volatilities, quasi-analytical 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. The model has a very good fit to an extensive panel data set of interest rates, swaptions and caps. In particular, the model matches the implied cap skews and the dynamics of implied volatilities. The model also performs well in forecasting interest rates and derivatives.

Anders B. Trolle

Copenhagen Business School

Solbjerg Plads 3, A5

DK-2000 Frederiksberg

DENMARK

abt.fi@cbs.dk

Eduardo S. Schwartz

UCLA Anderson School of Management

110 Westwood Plaza

Los Angeles, CA 90095-1481

and NBER

eduardo.schwartz@anderson.ucla.edu

1 Introduction

In this paper we develop a stochastic volatility multi-factor model of the term structure of interest rates based on the Heath, Jarrow, and Morton (1992) (HJM, henceforth) framework. The model has N factors driving the forward rate curve with each factor exhibiting stochastic volatility. The model allows for hump-shaped innovations to the forward rate curve. It also allows for correlations between innovations to forward rates and stochastic volatility which implies that the model has $N \times 2$ factors driving interest rate derivatives, except if correlations are perfect. The model has 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. In our model the dynamics of the forward rate curve under the risk neutral measure can be described in terms of a finite number of state variables which jointly follow an affine diffusion process. This facilitates pricing of more complex interest rate derivatives by Monte Carlo simulations. We apply the flexible “extended affine” market price of risk specification developed by Cheredito, Filipovic, and Kimmel (2003). This implies that the state vector also follows an affine diffusion process under the actual measure which facilitates econometric estimation.

We estimate the model for $N = 1, 2$ and 3 using an extensive panel data set consisting of 7 years 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 data set, 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 contains a number of contributions. Firstly, 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 which show that three factors are necessary to capture the variation in the term structure (see e.g. Litterman and Scheinkman (1991)), and that a number of additional factors unrelated to the term structure are needed to explain the variation in ATMF swaptions (Heidari and Wu (2003)), ATMF caps (Collin-Dufresne and Goldstein (2002a)) and non-ATMF caps (Li and Zhao (2006)). It is also consistent with Andersen and Benzoni (2005) who show that realized interest rate volatility is driven by

multiple unspanned factors.¹

Secondly, we address the relative valuation of swaptions and caps by re-estimating the $N=3$ model separately on swaptions and caps, and pricing caps and swaptions out-of-sample. We find that, according to our model, swaptions were mostly undervalued relative to caps during the first third of the sample. However, since then swaption and cap prices appear largely consistent with each other.

Thirdly, we highlight the importance of incorporating correlations between innovations to the term structure and the volatility factors. In the data we observe downward sloping cap skews in terms of log-normal implied volatilities with low strike, in-the-money caps trading at higher log-normal implied volatilities than high strike, out-of-the-money caps. We also observe that changes in log-normal implied volatilities are moderately negatively correlated with changes in the underlying forward rates, while changes in normal implied volatilities for swaptions and caps are moderately positively correlated with changes in the underlying forward rates.² 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. The fact that the correlation parameters which are consistent with implied volatilities across moneyness are also consistent with the dynamics of implied volatilities across time provides strong support for a correlation-based explanation for the implied cap skews.

Fourthly, as part of the estimation we obtain estimates of the market prices of risk associated with both the term structure factors and the volatility factors. This in turn facilitates use of the model for forecasting interest rates and interest rate derivatives in a consistent fashion. We show that our model performs well in this respect, especially at longer horizons. Crucial to this positive result is our use of the “extended affine” market price of risk specification. To our knowledge, this is the first paper that documents the ability of dynamic term structure

¹They analyze the US Treasury market, but their analysis most likely also holds for the LIBOR and swap markets due to the very high correlation between Treasury rates and LIBOR/swap rates of similar maturities.

²In this paper the term “log-normal implied volatility” is the volatility parameter which, plugged into the log-normal (or Black (1976)) pricing formula, matches a given price. The term “normal implied volatility” is the volatility parameter which, 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 log-normal implied volatility and $F(t, T)$ is the underlying forward rate. For non-ATMF swaptions and caplets the relation is approximately given by $\sigma_N = \sigma_{LN} \sqrt{F(t, T)K}$, where K is the the strike. See Blyth and Uglum (1999) and Zhou (2003).

models to forecast interest rate derivatives.³

Our model is related to the stochastic volatility LIBOR market models of Han (2004) and Jarrow, Li, and Zhao (2004). Han (2004) estimates his model on swaptions data, while Jarrow, Li, and Zhao (2004) estimate their model on cap skew data. In their models, conditional on the volatility factors, forward LIBOR rates are log-normally distributed, and forward swap rates are approximately log-normally distributed (under the appropriate forward measures). In contrast, in our model, conditional on the volatility factors, forward LIBOR and swap rates are approximately normally distributed (under the appropriate forward measures). More importantly, they impose zero correlations between innovations to forward LIBOR rates and volatility factors. The zero-correlation assumption implies that the forward LIBOR rate distributions have fatter tails than the log-normal distribution, and their models predict implied volatility smiles rather than the implied volatility skews that we observe.⁴ To capture the implied skews, Jarrow, Li, and Zhao (2004) add jumps with large negative mean jump sizes to the forward rate processes. The problem with the jump-based explanation for the implied cap skews, however, is that in the data we do not observe jumps even close to the magnitude necessary to fit the skews.⁵ Furthermore, the zero-correlation assumption implies that their models counter-factually predict that changes in log-normal implied volatilities are close to uncorrelated with changes in the underlying forward rates.⁶ It might seem logical, then, to extend the stochastic volatility LIBOR market model to non-zero correlations between innovations to forward LIBOR rates and stochastic volatility. Unfortunately, such a model is

³Cheredito, Filipovic, and Kimmel (2003) show that stochastic volatility term structure models with “extended affine” market price of risk specifications performs well in terms of forecasting interest rates. Collin-Dufresne, Goldstein, and Jones (2003) show that they also perform well in terms of forecasting realized volatilities. Neither study discuss interest rate derivatives.

⁴In 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 log-normal implied skews – steeper than observed in the data.

⁵Jarrow, Li, and Zhao (2004) argue that this could be due to extremely large jump risk premia, but the size of these risk premia seems implausible.

⁶In Han (2004)’s analysis this problem with the zero-correlation assumption is not apparent since his data set covers a period where interest rates were fairly stable. In contrast, Jarrow, Li, and Zhao (2004) report that the state variable that drives most of the stochastic volatility is strongly negatively correlated with interest rates despite the zero correlation assumption in their model. Note that in our model, a zero correlation assumption would imply that changes in normal implied volatilities were close to uncorrelated with changes in the underlying forward rates.

intractable.⁷ The ease with which we can incorporate non-zero correlations is one reason we prefer to work with instantaneous forward rates within the HJM framework. Another reason is the ability to obtain a finite dimensional affine model of the evolution of the forward rate curve.

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 non-zero correlation between innovations to forward rates and stochastic volatility.⁸

Other papers which use interest rate derivatives for estimating dynamic term structure models include Umantsev (2001), who use swaptions, and Bikbov and Chernov (2004), who use options on Eurodollar futures. These papers estimate traditional three-factor affine models which do not have sufficient flexibility to match the extensive data set used in this paper. Furthermore, in these models it is very difficult to generate “unspanned stochastic volatility” which arises naturally in our model.⁹

The paper is structured as follows. Section 2 sets up a general stochastic volatility term structure model. Section 3 discusses the data and the estimation procedure. Section 4 discusses the models’ fit to the term structure and derivatives and the relative valuation of swaptions and caps. Section 5 highlights the role of the correlation parameters in matching the implied cap skews and the dynamics of implied volatilities. Section 6 discusses the models’ forecasts of the term structure and derivatives. Section 7 concludes.

⁷The reason why non-zero 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 and the approximations necessary to retain analytical tractability, even with non-zero correlation.

⁸Andersen and Brotherton-Ratcliffe (2005) develop a CEV-type LIBOR market model with stochastic volatility that also has zero correlations between innovations to forward rates and volatility factors. 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 data set of interest rate derivatives.

⁹See 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”.

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

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

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

$i = 1, \dots, N$, where $W_i^Q(t)$ and $Z_i^Q(t)$ denote independent standard Wiener processes under the risk-neutral measure Q . The model differs from traditional HJM models by incorporating stochastic volatility, which is imperfectly correlated with forward rates. The model has N factors driving the forward rate curve and $N \times 2$ factors driving interest rate derivatives, and the model features some degree of “unspanned stochastic volatility” by construction, except in the special cases $\rho = -1$ or $\rho = 1$. 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 derivative literature.

Heath, Jarrow, and Morton (1992) show that absence of arbitrage implies that the drift term in (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 the risk-neutral probability measure is completely determined by the initial forward rate curve, the forward rate volatility functions and the volatility state variables.

To keep the model flexible yet analytically tractable, we specify the forward rate volatility functions $\sigma_{f,i}(t, T)$ in the following time-homogeneous way

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

This 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.

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 one to achieve identification.

In the following proposition we show that the dynamics of the forward curve can be described in terms of a finite number of state variables which jointly follow an affine diffusion process. This extends the results in Ritchkin and Chuang (1999) and Chiarella and Kwon (2003) to a stochastic volatility setting.

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

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

where

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

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

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

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

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

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

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

and the state variables evolve according to

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

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

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

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

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

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

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

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

Proof: See Appendix A. ■

The dynamics of the forward curve depends on $N \times 8$ state variables but there are only $N \times 2$ sources of risk, as there are no stochastic terms in the $\phi_{1,i}(t), \dots, \phi_{6,i}(t)$ processes. These are “auxiliary” state variables which simply reflect the path information of $x_i(t)$ and $v_i(t)$. 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 forward rate curve. In section 3 we make the model time-homogeneous 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

$$\begin{aligned} P(t, T) &\equiv \exp \left\{ - \int_t^T f(t, u) du \right\} \\ &= \frac{P(0, T)}{P(0, t)} \exp \left\{ \sum_{i=1}^N B_{x_i}(T-t)x_i(t) + \sum_{i=1}^N \sum_{j=1}^6 B_{\phi_{j,i}}(T-t)\phi_{j,i}(t) \right\} \end{aligned} \quad (20)$$

where

$$B_{x_i}(\tau) = \frac{\alpha_{1i}}{\gamma_i} \left(\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) (e^{-\gamma_i \tau} - 1) + \tau e^{-\gamma_i \tau} \right) \quad (21)$$

$$B_{\phi_{1,i}}(\tau) = \frac{\alpha_{1i}}{\gamma_i} (e^{-\gamma_i \tau} - 1) \quad (22)$$

$$B_{\phi_{2,i}}(\tau) = \left(\frac{\alpha_{1i}}{\gamma_i} \right)^2 \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) \left(\left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) (e^{-\gamma_i \tau} - 1) + \tau e^{-\gamma_i \tau} \right) \quad (23)$$

$$B_{\phi_{3,i}}(\tau) = -\frac{\alpha_{1i}}{\gamma_i^2} \left(\left(\frac{\alpha_{1i}}{2\gamma^2} + \frac{\alpha_{0i}}{\gamma} + \frac{\alpha_{0i}^2}{2\alpha_{1i}} \right) (e^{-2\gamma\tau} - 1) + \left(\frac{\alpha_{1i}}{\gamma} + \alpha_{0i} \right) \tau e^{-2\gamma\tau} + \frac{\alpha_{1i}}{2} \tau^2 e^{-2\gamma\tau} \right) \quad (24)$$

$$B_{\phi_{4,i}}(\tau) = \left(\frac{\alpha_{1i}}{\gamma_i} \right)^2 \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) (e^{-\gamma_i \tau} - 1) \quad (25)$$

$$B_{\phi_{5,i}}(\tau) = -\frac{\alpha_{1i}}{\gamma_i^2} \left(\left(\frac{\alpha_{1i}}{\gamma} + \alpha_{0i} \right) (e^{-2\gamma\tau} - 1) + \alpha_{1i} \tau e^{-2\gamma\tau} \right) \quad (26)$$

$$B_{\phi_{6,i}}(\tau) = -\frac{1}{2} \left(\frac{\alpha_{1i}}{\gamma_i} \right)^2 (e^{-2\gamma_i \tau} - 1). \quad (27)$$

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

$$\frac{dP(t, T)}{P(t, T)} = r(t)dt + \sum_{i=1}^N B_{x_i}(T-t) \sqrt{v_i(t)} dW_i^Q(t). \quad (28)$$

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

$$\psi(u, t, T_0, T_1) = E_t^Q \left[e^{-\int_t^{T_0} r_s ds} e^{u \log(P(T_0, T_1))} \right]. \quad (29)$$

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

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

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

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 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 $i = \sqrt{-1}$ and $G_{a,b}(y)$ is defined as

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

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 multi-factor 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 A 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 generated by our model.¹¹

Applying Ito's Lemma to the forward swap rate (see (72) in Appendix A) and switching

¹⁰Other approximation schemes have been developed by Collin-Dufresne and Goldstein (2002b), Singleton and Umantsev (2002) and Schrage and Pelsser (2005). However, these tend to be slower than the stochastic duration approach and hence not well suited for this paper, where a very large number of swaptions needs to be computed for each evaluation of the likelihood function.

¹¹To 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.

to the forward swap measure under which forward swap rates are martingales (see Jamshidian (1997)), we obtain

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

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

$$\begin{aligned} dv_i(t) = & \left(\kappa_i(\theta_i - v_i(t)) + v_i(t) \sigma_i \rho_i \nu \sum_{j=m+1}^n \xi_j(t) B_{x_i}(T_j - t) \right) dt \\ & + \sigma_i \sqrt{v_i(t)} \left(\rho_i dW_i^{Q^{T_m, T_n}}(t) + \sqrt{1 - \rho_i^2} dZ_i^{Q^{T_m, T_n}}(t) \right) \end{aligned} \quad (36)$$

where $\xi_j(t) = \frac{P(t, T_j)}{PVBP(t)}$. $W_i^{Q^{T_m, T_n}}(t)$ and $Z_i^{Q^{T_m, T_n}}(t)$ 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 $\zeta_j(t)$ terms are stochastic. Also, the process of $v_i(t)$ is non-affine under the forward swap measure due to the $\xi_j(t)$ terms. However, we can obtain an approximate and affine expression for the dynamics of the forward swap rate by replacing $\zeta_j(t)$ and $\xi_j(t)$ with their time t expected values, which are simply their time t values since these are martingales under the forward swap measure.¹² Therefore, 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) log-normally distributed.

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

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

¹²Due to the fact that $PVBP(t)$, which is the numeraire associated with the forward swap measure, appears in the denominators of these terms. A similar approach is followed by Schrage and Pelsser (2005) in a general affine model. They argue that the approximation is very accurate since $\zeta_j(t)$ and $\xi_j(t)$ typically have low variances.

where

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

An approximate price of the $(T_m - t)$ -into- $(T_n - T_m)$ swaption (i.e. an option expiring at T_m on a swap for the period T_m to T_n) can be obtained by inserting (38) in the normal swaption pricing formula. Monte Carlo evidence (not reported) shows this to be reasonably accurate for ATMF swaptions.¹³ Therefore we can view $\sigma_N(t, T_m, T_n)$ as a reasonably accurate expression of the normal implied ATMF swaption volatility generated by our model. The corresponding log-normal implied ATMF swaption volatility is approximately given by

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

These expressions yield several insights. Firstly, (38) and (39) directly link the volatility state variables in our model to the ATMF normal and log-normal implied volatilities. A positive $v_i(t)$ -shock naturally increases normal and log-normal implied volatility. 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 swaptions¹⁴ 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 options will be larger for the more persistent volatility state variables.

Secondly, 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 almost exclusively driven by the volatility factors, while the log-normal implied volatility surface is driven by both the term structure and volatility factors.

Thirdly, and related, without correlations between innovations to the term structure and volatility factors, the model predicts that changes in normal implied volatilities are close

¹³For $N = 3$ and typical parameter estimates reported in Section 4, the pricing errors range from -2 percent to 3 percent of the true price. Note that this way of pricing swaptions is extremely fast requiring only a single numerical integration. Therefore, it can be used in the initial stages of an optimization procedure to obtain a set of parameter estimates which can subsequently be refined with a more accurate approximation such as the stochastic duration approach.

¹⁴Where the expectation is taken under the forward swap measure.

to uncorrelated with changes in the underlying forward rates, while changes in log-normal implied volatilities are quite strongly negatively correlated with changes in the underlying forward rates. However, with positive correlation parameters the model predicts positive (less negative) correlations between normal (log-normal) 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 probability measure P which are obtained by specifying the market prices of risk, $\Lambda_{W,i}$ and $\Lambda_{Z,i}$ that link the Wiener processes under Q and P through

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

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

We apply the “extended affine” market price of risk specification suggested by Cheredito, Filipovic, and Kimmel (2003) 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

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

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

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

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

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

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 $\lambda_{Z,i0} = 0$.¹⁵ 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 measures. 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 or P . Therefore, we have to impose the following boundary non-attainment conditions¹⁶

$$2\kappa_i \theta_i \geq \sigma_i^2 \quad (46)$$

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

3 Estimation approach

3.1 Data

Our data set consists of weekly observations of LIBOR/swap term structures and log-normal implied ATMF swaption and cap volatilities from August 21, 1998 (i.e. just prior to the LTCM crisis) to July 8, 2005. In the latter half of the the sample from January 4, 2002 to July 8, 2005 we also have weekly observations on the log-normal implied cap skews. All observations are closing mid-quotes on Fridays and are obtained from Bloomberg.¹⁷

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

¹⁵Strictly speaking, in our setting the “essentially affine” specification coincides with the “completely affine” specification. However we could allow $dx_i(t) = \dots dt + \sqrt{\epsilon + v_i(t)} dW_i^Q(t)$, 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 (2003) for more on the this.

¹⁶Intuitively, if $v_i(t)$ were zero, we would have an infinite market price of risk despite zero volatility, representing an arbitrage opportunity. The boundary non-attainment conditions ensure that the market price of risk stays finite, although they can become arbitrarily large. The boundary non-attainment conditions must be satisfied under both P and Q for the measures to be equivalent. See Cheredito, Filipovic, and Kimmel (2003) for a further discussion.

¹⁷Note 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.

The swaptions have underlying swap maturities of 1yr, 2yr, 3yr, 5yr, 7yr and 10yr (called “tenors”) and option maturities of 1mth, 3mth, 6mth, 1yr, 2yr, 3yr and 5yr, 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 1yr, 2yr, 3yr, 4yr, 5yr, 7yr and 10yr. The strikes on the ATMF caps are the swap rates on the swaps with payments that corresponds 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 percent. 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 cap data, while Figure 4 displays the cap skew data. The missing data in the time series of skews for the 1yr and 2yr caps is due to the fact that very low interest rates have made a full skew unavailable in some periods (we refrain from extrapolating outside the range of implied volatilities that are available and only use 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) parametrization

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

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 log-normal (or Black (1976)) pricing formulas.

3.2 The Kalman filter

We estimate the model with the (extended) Kalman filter.¹⁸ This involves writing the model in state space form, which consists of a measurement equation and a transition equation. The

¹⁸Duffee 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

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

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

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

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

The y_t -vector consist of the LIBOR/swap term structure and the derivatives prices. The LIBOR and swap rates are non-linearly related to $x_1(t), \dots, x_N(t)$ and $\phi_{1,1}(t), \dots, \phi_{6,N}(t)$ through (20). For estimation we replace $f(0, T)$ with φ in (5) and $\frac{P(0, T)}{P(0, t)}$ with $\exp\{-\varphi(T - t)\}$ in (20) and estimate φ as part of the estimation procedure. This reduces the model to a time-homogeneous affine term structure model where φ is the infinite-maturity forward rate.¹⁹

The derivatives prices are non-linearly related to $v_1(t), \dots, v_N(t)$ through (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.²⁰ Since derivatives prices vary strongly across maturities of the options, maturities of the underlying LIBOR or swap rates as well as moneyness, we divide derivatives prices by their vegas, i.e. their sensitivities to variations in log-normal volatilities. With this scaling, the derivatives prices have comparable magnitudes.²¹

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 the more complex EMM or SML procedures would be extremely time-consuming, if not impossible.

¹⁹A similar approach is taken by de Jong and Santa-Clara (1999) in their estimation of HJM models.

²⁰When 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.

²¹This is very similar to fitting the model to log-normal 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.

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), $i = 1, \dots, N$,

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

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 (49) and (51), we need to linearize the h -function in (49) and make the assumption that the disturbance term w_t in (51) is Gaussian. With these modifications we can apply the (extended) Kalman filter to (49) and (51) and in the process obtain the likelihood function (for completeness the (extended) Kalman filter recursions are stated in Appendix B. Harvey (1989) and Hamilton (1994) are classic references).²²

3.3 Numerical issues

The log-likelihood 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 20 integration points and truncating the integral at 8000. For the model with $N=3$ estimated on the entire data set, each evaluation of the likelihood function requires calculating 60,480 swaption prices and 514,336 caplet prices,²³ underscoring the need for fast pricing

²²The use of a Gaussian distribution to approximate the true distribution of w_{t+1} makes this a QML procedure. While QML estimation has been shown to be consistent in many settings, it is in fact not consistent when used in conjunction with the Kalman filter (see. e.g. Lund (1997)). However, the Monte Carlo studies in Duffee and Stanton (2004) and in several other papers show the inconsistency problem to be of minor importance in the context of term structure estimation.

²³In 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 (82) in Appendix B is computed

routines.

4 Estimation results

4.1 Parameter estimates

We start by estimating our model for $N=1$ (i.e. one term structure factor and one volatility factor), $N=2$ (i.e. two term structure factors and two volatility factors) and $N=3$ (i.e. three term structure factors and three volatility factors) on the entire data set. We also re-estimate the model for $N=3$ on the swaption and cap data separately and price caps and swaptions out-of-sample to address the relative valuation of swaptions and caps.

The five sets of parameter estimates are given in Table 1 and 2. For all the models the estimates of $\alpha_{0,i}$, $\alpha_{1,i}$ and γ_i imply that all the forward rate volatility functions are hump shaped, with $\sigma_{f,1}(t, T)$ affecting the entire forward rate curve, $\sigma_{f,2}(t, T)$ affecting only the short end of the curve and $\sigma_{f,3}(t, T)$ affecting mainly the intermediate part of the curve. This is consistent with principal component analyses of the yield curve, which find that the three dominant factors are the “level”, “slope” and “curvature” factors. Panel A in Figure 5 displays the forward rate volatility functions in case of the $N=3$ swaption and cap model.

For all the models the first volatility factor is more persistent than the second volatility factor under the risk-neutral measure. Interestingly, the third volatility factor is the least persistent for the $N=3$ swaption and cap model and swaption model but the most persistent for the $N=3$ cap model. The volatility factors are always less persistent under the actual measure than under the risk-neutral measure. Panel B in Figure 5 displays the volatility state variables in case of the $N=3$ swaption and cap model.

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

In all the estimations, the boundary non-attainment condition is binding for all the volatil-

numerically so we need to reprice the swaptions and caplets for small perturbations of $v_1(t)$, $v_2(t)$ and $v_3(t)$.

ity processes under Q but not under P . We have re-estimated the models with the “completely affine” market price of risk specification, which does not impose the boundary non-attainment condition. 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 model’s 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 the results reported by Cheredito, Filipovic, and Kimmel (2003) in the context of term structure estimation.

For all the models, the correlation parameters are positive, and at least one correlation parameter is significantly different from zero. This is consistent with the preliminary observations made earlier in the paper that non-zero correlations were crucial for matching the implied cap skews and the dynamics of implied volatilities. We return to the role of the correlation parameters in Section 5.

In general the Q -parameters are much more precisely estimated than the P -parameters, which is not surprising given the relatively short time-series. Particularly the P -parameters in the dynamics of the $x_i(t)$ state variables are very imprecisely estimated.

4.2 Interpretation of factors

As discussed in Section 2.3, variations in the normal implied volatilities are almost exclusively driven by variations in the volatility factors, while variations in the log-normal implied volatilities are driven by variations in both the term structure and volatility factors. To provide intuition about the dynamics of the model, Figure 6 shows the impact on the log-normal implied swaption volatility surface from one-standard deviation shocks to $x_1(t)$, $x_2(t)$, $x_3(t)$, $v_1(t)$, $v_2(t)$ and $v_3(t)$ in case of the $N=3$ swaption and cap model, assuming that the zero-coupon curve and $v_1(t)$, $v_2(t)$ and $v_3(t)$ are initially equal to their sample averages.

Our model allows for a rich set of shocks to the implied volatility surface. The effect of a shock to a term structure factor depends on the impact on the forward swap rates underlying the swaptions. A $x_1(t)$ -shock decreases the entire implied volatility surface with a slight inverse hump shaped impact along the option dimension for short swaps and along the swap dimension for short options. A $x_2(t)$ -shock mainly decreases the implied volatility of short options on short swaps, and the effect quickly diminishes with the length of the option and the length of the swap. Compared with a $x_1(t)$ -shock, a $x_3(t)$ -shock has a more pronounced inverse hump

shaped impact along the option dimension for short swaps and along the swap dimension for short options.

The effect of a shock to a volatility factor depends on the impact on the instantaneous volatilities of the forward swap rates underlying the swaptions and the persistence of the volatility shock. A $v_1(t)$ -shock is moderately persistent and increases the instantaneous volatility of all forward swap rates. The result is an increase in the entire implied volatility surface, with the implied volatilities of short options on longer swaps increasing the most. A $v_2(t)$ -shock is less persistent and mainly affects the instantaneous volatility of short swaps with short forward starts. The result is that only short options on short swaps are affected. A $v_3(t)$ -shock is very persistent and mainly affects the instantaneous volatility of intermediate forward swaps (short swaps with intermediate forward starts or intermediate swaps with short forward starts). The result is a hump shaped impact along the option dimension for short swaps along the swap dimension for short options.

For completeness Figure 7, Panel A, B and C shows the impact on the log-normal implied ATMF cap volatilities from one-standard deviation shocks to $x_1(t)$, $x_2(t)$, $x_3(t)$, $v_1(t)$, $v_2(t)$ and $v_3(t)$.²⁴ These are similar to the impact on implied volatilities of swaptions with short tenors.

4.3 Overall comparisons of models

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. By taking the square root of the average of the squared pricing errors at each date, we construct time-series of root mean squared errors (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 signifi-

²⁴We discuss Panel D, E and F below.

cantly negative mean implies that model two has a significantly better fit to caps than model one (according to the RMSE criterion). When computing the t -statistics we use Newey and West (1987) standard errors with 12 lags to correct for heteroscedasticity and autocorrelation.²⁵

Table 3 displays the average RMSEs for 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. For the swaption and cap models, the fit improves with the number of factors and the decreases in average RMSEs between the $N=2$ and $N=1$ models and between the $N=3$ and $N=2$ models 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 $N=3$ swaption model has a superior fit to swaptions, but an inferior (out-of-sample) fit to caps than the $N=3$ swaption and cap model. The converse holds for the $N=3$ cap model, which has a superior fit to caps but an inferior (out-of-sample) fit to swaptions compared to the $N=3$ swaption and cap model. 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.

4.4 In-sample fit

We now take a closer look at the fit of the $N=3$ models. Figure 8 displays the time series of the RMSEs of LIBOR/swap rates, ATMF swaptions, ATMF caps and non-ATMF caps for the $N=3$ swaption model (dotted lines) and the $N=3$ cap model (solid lines). 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 the $N=3$ models. In this section we focus on the in-sample fit while in the next section we

²⁵The results are robust to variations in the lag length.

focus on the out-of-sample fit.

The in-sample swaption RMSE (dotted line, Panel B) reaches about 15 percent in September 1998 during the LTCM crisis. Longstaff, Santa-Clara, and Schwartz (2001) and Han (2004) also report a significant increase in swaption pricing errors during this period. The in-sample swaption RMSE reaches about ten percent 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 two to six percent. The RMSE is comparable to that reported by Han (2004) for his preferred model with four term structure factors and three volatility factors during the sample period which overlaps with ours. Note, however, that we include a larger number of swaptions than his study. In particular, our data set includes 1mth and 3mth options and 10yr underlying swaps, which are not present in his data set. And it is precisely these swaptions on the “edges” of the volatility surface that are the most difficult to fit.

The in-sample ATMF cap RMSE (solid line, Panel C) also spikes in September 1998. Otherwise it mostly fluctuates between one and two percent. The in-sample non-ATMF 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 significantly lower than for the preferred model in Jarrow, Li, and Zhao (2004) with three term structure factors, three volatility factors and jumps during the sample period which overlaps with ours (they report that the RMSE fluctuates around five percent during this period).

For the $N=3$ swaption and cap model the average swaption errors range from -3.41 to 3.39 percent, the average ATMF cap errors range from 3.08 and 0.12 percent and the average non-ATMF cap errors range from -4.17 to 4.15 percent. Quite a few of the pricing errors are statistically significant.

For the $N=3$ swaption model the range of average swaption errors narrows to -2.84 to 2.52 percent. 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 re-calibrated at every date, lie in a range from -5.37 to 5.62 percent.

For the $N=3$ cap model, the range of average pricing errors narrows to -0.01 to 0.40 percent for ATMF caps and -1.51 to 1.59 percent for non-ATMF caps. To put these numbers into perspective, the mean pricing errors reported by Jarrow, Li, and Zhao (2004) for their

preferred model estimated on cap skew data, although not for exactly the same sample period, lie in a range from -6.88 percent to 7.13 percent. Note also that for the $N=3$ cap 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 one and 10 basis points, and the average errors are within a few basis points with no apparent differences between the models.

Figures 9 – 11 displays the average of the actual and fitted term structure, normal implied swaption volatility surface and normal implied ATMF cap term structure, respectively, in case of the $N=3$ swaption and cap model.²⁶ The actual and fitted term structure is indistinguishable from each other, and the fit to implied volatilities are also very close.

As discussed by Dai and Singleton (2002), the fitted data depends not only on the properties of a model but also on the properties of the historical data used for estimation. Therefore, comparing 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 1000 samples of term structures and implied volatilities from the $N=3$ swaption and cap model. Each sample consists of 360 weekly observations similar to our original data set. From these we obtain the small-sample distributions of the average term structure, average swaption volatility surface and average ATMF cap volatility term structure generated by the model. The means and 95 percent confidence intervals of these distributions are also displayed in Figures 9 – 11. The means of the small-sample distributions are close to the means of the actual data, and the means of the actual data are certainly well within the 95 percent confidence intervals of the small-sample distributions. This underscores the very good fit of our model.

4.5 Out-of-sample fit – the relative valuation of caps and swaptions

We now turn to the out-of-sample fit to swaptions and caps, i.e. the fit to caps for the $N=3$ swaption model and the fit to swaptions for the $N=3$ cap model. We are particularly interested in whether caps and swaption are priced consistently with each other. In Figure 8 the out-of-sample swaptions RMSE (solid line, Panel B) and out-of-sample cap RMSEs (dotted

²⁶We display the simulated derivatives data in terms of normal rather than log-normal implied volatilities since the normal implied volatilities exhibit a more pronounced hump-shape that most dynamic term structure models have difficulties matching.

lines, Panel C and D) are larger most of the time than their in-sample counterparts. This is particularly the case in the first third of the sample.

For the $N=3$ swaption model, the average cap errors 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 $N=3$ cap model, the average swaption errors are significantly positive for swaptions with underlying swap maturities of 1yr, 2yr and 3yr (except for the 1mth-into-1yr swaption) and significantly negative for swaptions with underlying swap maturities of 7yr and 10yr. However, the out-of-sample results are probably most reliable for swaptions with combined swap and option maturity not exceeding ten 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 percent. 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 $N=3$ swaption 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 (2004) finds little mis-valuation 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 12 shows the average (out-of-sample) swaption valuation errors according to the $N=3$ cap model (the solid line) and the average (out-of-sample) cap valuation errors according to the $N=3$ swaption 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.²⁷ However, since then there appear to be little systematic mis-valuation in the aggregate between swaptions and caps.

²⁷Han (2004) 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.

5 The role of the interest rate – volatility correlations

5.1 Matching the implied cap skews

As discussed in Section 2.3, conditional on the volatility factors, forward LIBOR and swap rates are approximately normally distributed in our model (under the appropriate forward measure). Suppose the correlation between innovations to the forward 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 log-normal implied volatilities, with ITM caps trading at higher log-normal implied volatilities 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 depend on the correlation parameters. To illustrate this, Figure 7, Panel D, E and F shows the derivatives of the differences between non-ATMF and ATMF log-normal implied volatilities with respect to the correlation parameters. The results are for the $N=3$ swaption and cap model, assuming that the zero-coupon curve and $v_1(t)$, $v_2(t)$ and $v_3(t)$ are initially equal to their sample averages. In all cases, increasing the correlation parameters decreases the log-normal implied volatilities of ITM caps relative to OTM caps which decreases the steepness of the implied cap skews. Considering Panel A, B and C it is not surprising that ρ_1 mainly affects the implied skews of long-term caps, ρ_2 mainly affects the implied skews of short-term caps while ρ_3 has the largest effect on implied skews of intermediate-maturity caps.

Figure 13 shows the average fit to the actual implied cap skews (solid lines) for the $N=3$ cap model (dash-dotted lines) and the $N=3$ cap model re-estimated with the correlation parameters restricted to zero (dotted lines). We see that with zero correlations the model produces implied skews that are too steep on average. In contrast, with non-zero correlations the model has an almost perfect fit to the implied skews on average. The picture is basically the same for the $N=3$ swaption and cap model (not shown) although this model tends to slightly overestimate the average steepness of the implied skews, particularly for caps of intermediate maturities.

5.2 Matching the dynamics of implied volatilities

Figure 14, Panel A shows the correlations between changes in log-normal implied swaption volatilities and changes in the underlying forward swap rates. For all the swaptions the cor-

relations are negative, more so for longer swaptions. Panel D shows the correlations using normal rather than log-normal implied swaption volatilities. In this case, all the correlations are positive.²⁸

As discussed in Section 2.3, our model with zero correlation between innovations to the forward curve and volatilities would predict that changes in normal implied swaption volatilities were close to uncorrelated with changes in the underlying forward rates, and that changes in log-normal implied swaption volatilities were quite strongly negatively correlated with changes in the underlying forward rates. However, with non-zero correlations the model should be able to match the actual dynamics of implied volatilities more closely.

To see if this is the case for the $N=3$ swaption and cap model we use the simulated samples discussed in Section 4.4. For each sample, we compute correlations between changes in normal and log-normal 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. Panel B and E in Figure 14 display the means of these distributions, while Panel C and F display the 95 percent 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 percent confidence bands. The model has a reasonable fit to the log-normal 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 percent confidence bands.

5.3 Statistical significance of the correlation parameters

The previous two subsections show that incorporating non-zero correlations has important practical implications for the ability of the model to match the implied cap skew and the dynamics of implied volatilities. Here we investigate if incorporating non-zero correlations is important from a statistical point of view. We re-estimate the $N=3$ models with the three correlation parameters restricted to zero. This yields restricted log-likelihood values of -33096.7, -18952.7 and -5523.3 for the 3SC, 3S and 3C models, respectively. The corresponding unrestricted log-likelihood values from Table 1 and 2 are -32887.5, -18947.9 and -3919.2, respectively. Therefore, the likelihood ratio test statistics of the zero-correlation restriction are 418.4,

²⁸These stylized facts are quite robust. For instance, computing the correlations using only the first half or the second half of the sample yield similar results. These stylized facts also hold for caps but to avoid an overload of figures we concentrate on swaptions in this section.

9.6 and 3208.2 for the 3SC, 3S and 3C models, respectively. These should be compared with the critical values of a $\chi^2(3)$ -distribution which are 7.81 at the five percent level and 11.34 at the one percent level. For the $N=3$ swaption and cap model and the $N=3$ cap model we strongly reject the zero-correlation restriction. However, for the $N=3$ swaption model we just barely reject the zero-correlation restriction at conventional levels.²⁹ The difference is that the former two models include cap skew data in the estimation which implies that the correlation parameters can be identified from both the variation in implied volatilities across both moneyness and across time. In the latter model cap skew data is not included and the correlation parameters can only be identified from the variation in implied volatilities across time. We can therefore conclude that the variation in implied volatilities across moneyness provides much stronger identification of the correlation parameters than the variation in implied volatilities across time.

6 Forecasting interest rates and derivatives

So far we have been concerned with the fit of the models, which mainly depend on the dynamics under the risk-neutral measure. In this section we turn to the forecasting performances of the models, which depend critically on the dynamics under the actual measure. We forecast the term structure and interest rate derivatives using our five estimated models. To evaluate the quality of the forecasts we also include the benchmarks that interest rates and derivatives prices follow a random walk, that is, the best forecast of future term structures and derivatives prices are today's term structure and derivatives prices.³⁰

We consider forecast horizons of 4, 8 and 16 weeks. For the non-ATMF caps, we exclude the 1yr caps, as there are too few observations to reliably evaluate forecasts, especially at longer horizons. All forecasts are in-sample. For the LIBOR and swap rates, we take the forecast

²⁹This is consistent with the fact that in Table 1 and 2 the standard errors on the correlation parameters are higher for the $N=3$ swaption model than for the other two models. The reason why the zero-correlation restriction is more strongly rejected for the $N=3$ cap model than the $N=3$ swaption and cap model is that the unrestricted $N=3$ cap model has a closer fit to the implied cap skews causing a larger decline in the likelihood value when imposing the zero-correlation restriction.

³⁰We could include more sophisticated benchmarks but the random walk is a standard benchmark in the forecasting literature. Since we are forecasting prices of newly issued derivatives, the random walk benchmark seems valid at shorter horizons despite the fact that it does imply positive probabilities of negative prices of non-negative contingent claims, which would present an arbitrage opportunity.

errors to be the difference between the predicted and actual interest rate changes. For the swaptions and caps, we take the forecast errors to be the difference between the predicted and actual percentage price changes. For each forecast horizon 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 forecasting RMSEs.

Table 9 displays the overall forecasting RMSEs for each of the five models at each of the three forecasting horizons, and Table 10 makes pairwise comparisons of forecasting performance using the Diebold and Mariano (1995) approach outlined in section 4.3.

For the swaption and cap models the forecasting performance generally improves with the number of factors with the decreases in average RMSEs largest between the $N=2$ and $N=1$ models. Even the $N=1$ swaption and cap model beats the random walk benchmarks for both interest rates and derivatives, and for the $N=3$ swaption and cap model the improvement over the random walk benchmarks is strongly significant for the term structure and ATMF caps.

The $N=3$ swaption model outperforms the $N=3$ swaption and cap model in terms of forecasting swaptions but produces worse (out-of-sample) forecasts of caps. The $N=3$ cap model outperforms the $N=3$ swaption and cap model in terms of forecasting ATMF caps but produces worse (out-of-sample) forecasts of swaptions.

We have also estimated the models with the “completely affine” market price of risk specification. These models produce significantly worse forecasts for derivatives, especially at longer horizons. The reason is that these models generally do not match the long run means of the volatility state variables under the actual measure. The added flexibility of the “extended affine” market price of risk specification therefore seems crucial for producing good forecasts of derivatives.

7 Conclusion

We have developed a flexible stochastic volatility multi-factor model of the term structure of interest rates. It features possibly hump-shaped innovations to the forward rate curve and correlations between innovations to forward rates and stochastic volatility (i.e. partially unspanned stochastic volatility). The model is highly tractable with quasi-analytical 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 data set of LIBOR and swap rates, ATMF swaptions, ATMF caps and, for the second half of the sample, non-ATMF caps (i.e. cap skews). With three term structure factors and three volatility factors the model has a very good fit to the data. Re-estimating the model on swaptions and caps separately and pricing caps and swaptions out-of-sample reveals that swaptions were mostly undervalued relative to caps during the first third of the sample (at least relative to our model). However, since then swaption and cap prices appear largely consistent with each other.

A key result is the ability of the model to match the implied cap skews and the dynamics of implied volatilities. This hinges on the correlations between innovations to forward rates and stochastic volatility. In relation to estimating the correlations we show that the variation in implied volatilities across moneyness provides much stronger identification of the correlation parameters than the variation in implied volatilities across time.

Finally, the model performs well in terms of forecasting interest rates and interest rate derivatives. Crucial to this result is the use of the flexible “extended affine” market price of risk specification.

Our model has many applications. Firstly, the ease with which the risk-neutral dynamics of the forward rate curve can be simulated makes it useful for pricing more complex interest rate derivatives by Monte Carlo simulations. 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.³¹

Secondly, the good forecasting performance of the model, combined with the ease with which the actual dynamics of the forward rate curve can be simulated, suggests that it will be useful in value-at-risk applications involving portfolios of interest rate derivatives.

Thirdly, the model can be extended to price commodity futures and options in a stochastic volatility HJM framework. These applications are the subject of ongoing research.

³¹Many 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.

Appendix A

Proof of Proposition 1

Given (4), (3) becomes

$$\begin{aligned} \mu_f(t, T) = & \sum_{i=1}^N v_i(t) \left[\frac{\alpha_{0i}\alpha_{1i}}{\gamma_i} \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) \left(e^{-\gamma_i(T-t)} - e^{-2\gamma_i(T-t)} \right) - \frac{\alpha_{0i}\alpha_{1i}}{\gamma_i} (T-t) e^{-2\gamma_i(T-t)} \right. \\ & \left. + \frac{\alpha_{1i}^2}{\gamma_i} \left(\frac{1}{\gamma_i} + \frac{\alpha_{0i}}{\alpha_{1i}} \right) (T-t) \left(e^{-\gamma_i(T-t)} - e^{-2\gamma_i(T-t)} \right) - \frac{\alpha_{1i}^2}{\gamma_i} (T-t)^2 e^{-2\gamma_i(T-t)} \right] \end{aligned} \quad (52)$$

Straightforward, if slightly tedious, calculations show that

$$f(t, T) = f(0, T) + \int_0^t \mu_f(s, T) ds + \sum_{i=1}^N \int_0^t \sigma_{f,i}(s, T) \sqrt{v_i(s)} dW_i^Q(s) \quad (53)$$

$$= f(0, T) + \sum_{i=1}^N \mathcal{B}_{x_i}(T-t) x_i(t) + \sum_{i=1}^N \sum_{j=1}^6 \mathcal{B}_{\phi_{j,i}}(T-t) \phi_{j,i}(t) \quad (54)$$

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

$$x_i(t) = \int_0^t \sqrt{v_i(s)} e^{-\gamma_i(t-s)} dW_i^Q(s) \quad (55)$$

$$\phi_{1,i}(t) = \int_0^t \sqrt{v_i(s)} (t-s) e^{-\gamma_i(t-s)} dW_i^Q(s) \quad (56)$$

$$\phi_{2,i}(t) = \int_0^t v_i(s) e^{-\gamma_i(t-s)} ds \quad (57)$$

$$\phi_{3,i}(t) = \int_0^t v_i(s) e^{-2\gamma_i(t-s)} ds \quad (58)$$

$$\phi_{4,i}(t) = \int_0^t v_i(s) (t-s) e^{-\gamma_i(t-s)} ds \quad (59)$$

$$\phi_{5,i}(t) = \int_0^t v_i(s) (t-s) e^{-2\gamma_i(t-s)} ds \quad (60)$$

$$\phi_{6,i}(t) = \int_0^t v_i(s) (t-s)^2 e^{-2\gamma_i(t-s)} ds \quad (61)$$

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

Proof of Proposition 2

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

$$e^{-\int_0^t r_s ds} \psi(u, t, T_0, T_1) = E_t^Q \left[e^{-\int_0^{T_0} r_s ds} \psi(u, T_0, T_0, T_1) \right]. \quad (62)$$

since

$$\psi(u, T_0, T_0, T_1) = e^{u \log(P(T_0, T_1))} \quad (63)$$

Therefore, the proof consist of showing that the process

$$\eta(t) \equiv e^{-\int_0^t r_s ds} \psi(u, t, T_0, T_1). \quad (64)$$

is a martingale under Q . To this end we conjecture that $\psi(u, t, T_0, T_1)$ is of the form (30). Applying Ito'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 (31)–(32). Furthermore, (63) 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 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

$$\begin{aligned} \mathcal{P}(t, T_0, T_1, K) &= E_t^Q \left[e^{-\int_t^{T_0} r(s) ds} (K - P(T_0, T_1)) \mathbf{1}_{P(T_0, T_1) < K} \right] \\ &= K E_t^Q \left[e^{-\int_t^{T_0} r(s) ds} \mathbf{1}_{\log(P(T_0, T_1)) < \log(K)} \right] - \\ &\quad E_t^Q \left[e^{-\int_t^{T_0} r(s) ds} e^{\log(P(T_0, T_1))} \mathbf{1}_{\log(P(T_0, T_1)) < \log(K)} \right] \\ &= K G_{0,1}(\log(K)) - G_{1,1}(\log(K)), \end{aligned} \quad (65)$$

where

$$G_{a,b}(y) = E_t^Q \left[e^{-\int_t^{T_0} r(s) ds} e^{a \log(P(T_0, T_1))} \mathbf{1}_{b \log(P(T_0, T_1)) < y} \right]. \quad (66)$$

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

$$\begin{aligned} \mathcal{G}_{a,b}(y) &= \int_{\mathbb{R}} e^{iuy} dG_{a,b}(y) \\ &= E_t^Q \left[e^{-\int_t^{T_0} r_s ds} e^{(a+iub) \log(P(T_0, T_1))} \right] \\ &= \psi(a + iub, t, T_0, T_1), \end{aligned} \quad (67)$$

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

$$G_{a,b}(y) = \frac{\psi(a, t, T_0, T_1)}{2} - \frac{1}{\pi} \int_0^\infty \frac{\text{Im} [\psi(a + iub, t, T_0, T_1) e^{-iuy}]}{u} du. \quad (68)$$

LIBOR and swap rates

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

$$L(t, T) = \frac{1 - P(t, T)}{(T - t)P(t, T)}. \quad (69)$$

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

$$L(t, T_1, T_2) = \frac{P(t, T_1) - P(t, T_2)}{(T_2 - T_1)P(t, T_2)}. \quad (70)$$

In the following, consider a period length ν and a set of dates $T_j = t + \nu j$, $j = 1, \dots, n$. The time- t swap rate for the period t to T_n and fixed-leg payments at T_1, \dots, T_n ³² is given by

$$S(t, T_n) = \frac{1 - P(t, T_n)}{\nu \sum_{j=1}^n P(t, T_j)}. \quad (71)$$

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

$$S(t, T_m, T_n) = \frac{P(t, T_m) - P(t, T_n)}{\nu \sum_{j=m+1}^n P(t, T_j)}. \quad (72)$$

Cap prices

A cap with payments at T_2, \dots, T_n consists of $n - 1$ caplets expiring at these dates.³³ 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

$$\chi = \nu (L(T_j - \nu, T_j) - K)^+ \quad (73)$$

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

$$\text{Cpl}(t, T_j - \nu, T_j, X_t, K) = (1 + \nu K) \mathcal{P} \left(t, T_j - \nu, T_j, \frac{1}{1 + \nu K} \right). \quad (74)$$

³²Market convention for USD swaps is semi-annual 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 practice, USD LIBOR derivatives (swaps and caps) 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.

³³Market 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 at time t .

The time- t price of a cap with strike R and payments at T_2, \dots, T_n is the sum of the prices of the $n - 1$ underlying caplets

$$\text{Cap}(t, T_n, X_t, R) = \sum_{j=2}^n \text{Cpl}(t, T_j - \nu, T_j, X_t, R). \quad (75)$$

When the strike R equals the the time- t forward swap rate for the period T_1 to T_n with fixed-leg payments at T_2, \dots, T_n the cap price equals the floor price and the cap and floor are at-the-money-forward.

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}, \dots, T_n and fixed rate K is given by

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

A payer swaption is an option to enter into a payer swap with 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

$$\chi = V(T_m, T_m, T_n)^+ = \left(1 - P(T_m, T_n) - K\nu \sum_{j=m+1}^n P(T_m, T_j) \right)^+. \quad (77)$$

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,

$$P^c(t) = \sum_{j=m+1}^n Y(T_j)P(t, T_j), \quad (78)$$

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

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

$$\sum_{i=1}^N v_i(t) B_{x_i}(D(t))^2 = \sum_{i=1}^N v_i(t) \left(\sum_{j=m+1}^n w_j B_{x_i}(T_j - t) \right)^2, \quad (79)$$

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 exist and is unique if $B_{x_i}(\tau)$ is uniformly decreasing, see Munk (1999).³⁴

Wei (1997) and Munk (1999) suggest approximating an option on a coupon bond option with an (scaled) option on a zero-coupon bond with maturity equal to the stochastic duration of the coupon bond. Let $\text{Swpn}(t, T_m, T_n, K)$ 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

$$\text{Swpn}(t, T_m, T_n, K) = \zeta \mathcal{P}(t, T_m, D(t), \zeta^{-1}), \quad (80)$$

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

Appendix B

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 (49) around $\hat{X}_{t|t-1}$ we obtain

$$y_t = (h(\hat{X}_{t|t-1}) - H'_t \hat{X}_{t|t-1}) + H'_t X_t + u_t, \quad u_t \sim \text{iid. } N(0, S), \quad (81)$$

where

$$H'_t = \left. \frac{\delta h(X_t)}{\delta X'_t} \right|_{X_t = \hat{X}_{t|t-1}}. \quad (82)$$

Assuming w_t in (51) is Gaussian we obtain

$$X_{t+1} = \Phi_0 + \Phi_X X_t + w_{t+1}, \quad w_{t+1} \sim \text{iid. } N(0, Q_t). \quad (83)$$

The Kalman filter applied to (81) and (83) yields

$$\hat{X}_{t+1|t} = \Phi_0 + \Phi_X \hat{X}_t \quad (84)$$

$$P_{t+1|t} = \Phi_X P_t \Phi'_X + Q_t \quad (85)$$

and

$$\hat{X}_{t+1} = \hat{X}_{t+1|t} + P_{t+1|t} H'_t F_t^{-1} \epsilon_t \quad (86)$$

$$P_{t+1} = P_{t+1|t} - P_{t+1|t} H'_t F_t^{-1} H_t P_{t+1|t}, \quad (87)$$

³⁴At 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.

where

$$\epsilon_t = y_{t+1} - h(\hat{X}_{t+1|t}) \quad (88)$$

$$F_t = H_t P_{t+1|t} H_t' + S. \quad (89)$$

The log-likelihood function is constructed from (88) and (89):

$$\log L = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^T \log |F_t| - \frac{1}{2} \sum_{i=1}^T \epsilon_t' F_t^{-1} \epsilon_t, \quad (90)$$

where T is the number of observations and N is the dimension of y_t . We follow standard practice in the literature and initialize the Kalman filter at the unconditional values of $\hat{X}_{t|t-1}$ and $P_{t|t-1}$.

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

Notes: 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 one.

Table 1: Parameter estimates

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

Notes: 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 one.

Table 2: Parameter estimates (cont.)

	Model				
	1SC	2SC	3SC	3S	3C
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

Notes: 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.

Table 3: In-sample fit

Model comparisons				
	2SC vs. 1SC	3SC vs. 2SC	3S vs. 3SC	3C vs. 3SC
Interest rates	-39.03*** (-9.07)	-5.36*** (-9.06)	0.05 (0.79)	-0.18 (-1.02)
ATMF swaptions	-4.08*** (-16.07)	-2.05*** (-13.43)	-0.53*** (-9.13)	7.96*** (9.55)
ATMF caps	-3.50*** (-6.70)	-0.20 (-0.67)	2.66*** (9.01)	-2.59*** (-5.13)
Non-ATMF caps	-1.62*** (-2.59)	-1.19*** (-4.42)	2.10*** (8.96)	-1.25*** (-4.98)

Notes: 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.

Table 4: Statistical comparisons of in-sample fit

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

Notes: 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.

Table 5: Summary statistics for LIBOR and swap rate errors

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

Notes: 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.

Table 6: Summary statistics for ATMF swaption valuation errors

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

Notes: 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.

Table 7: Summary statistics for ATMF cap valuation errors

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

Notes: 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.

Table 8: Summary statistics for non-ATMF cap valuation errors

		Model					
		1SC	2SC	3SC	3S	3C	RW
Interest rates	4wks	21.17	20.70	20.65	20.76	20.66	22.27
	8wks	32.50	31.29	31.38	31.67	31.35	34.83
	16wks	51.23	47.95	47.97	48.77	47.66	55.15
ATMF swaptions	4wks	7.72	7.75	7.77	7.75	8.01	7.72
	8wks	10.67	10.59	10.63	10.58	11.27	10.83
	16wks	14.00	13.54	13.58	13.46	14.79	14.56
ATMF caps	4wks	7.83	7.81	7.75	7.78	7.66	8.01
	8wks	10.69	10.53	10.34	10.40	10.08	11.13
	16wks	14.68	14.17	13.72	13.81	13.24	15.79
Non-ATMF caps	4wks	7.42	7.43	7.43	7.47	7.49	7.54
	8wks	10.27	10.19	10.18	10.25	10.27	10.50
	16wks	14.47	14.02	13.85	13.98	13.74	14.47

Notes: Mean of root mean squared forecasting errors. For interest rates the forecast errors are the differences between the predicted and actual interest rate changes. For swaptions and caps the forecast errors are the differences between the predicted and actual percentage price changes. “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, “3C” denotes the $N=3$ cap model, and “RW” denotes the random walk model for interest rates and derivatives prices. Interest rate forecast errors are measured in basis points while derivative forecast errors are measured in percentages.

Table 9: Forecasting performance

		Model comparisons				
		2SC vs.	3SC vs.	3S vs.	3C vs.	RW vs.
		1SC	2SC	3SC	3SC	3SC
Interest rates	4wks	-0.48 (-1.34)	-0.05 (-0.24)	0.11 (1.62)	0.01 (0.06)	1.62*** (3.26)
	8wks	-1.21 (-1.31)	0.08 (0.15)	0.29* (1.87)	-0.03 (-0.10)	3.45*** (2.73)
	16wks	-3.27 (-1.63)	0.01 (0.01)	0.80** (2.12)	-0.31 (-0.38)	7.18*** (2.75)
ATMF swaptions	4wks	0.03 (0.38)	0.03 (1.32)	-0.02 (-0.98)	0.24 (1.30)	-0.05 (-0.28)
	8wks	-0.08 (-0.56)	0.04 (0.99)	-0.05 (-1.46)	0.64 (1.31)	0.19 (0.45)
	16wks	-0.47* (-1.73)	0.04 (0.48)	-0.12** (-2.18)	1.21 (1.05)	0.98 (1.10)
ATMF caps	4wks	-0.02 (-0.24)	-0.05 (-1.38)	0.03 (0.93)	-0.09 (-1.13)	0.26* (1.92)
	8wks	-0.15 (-0.75)	-0.19** (-1.97)	0.06 (1.02)	-0.26 (-1.59)	0.78*** (2.69)
	16wks	-0.51 (-1.36)	-0.45** (-2.43)	0.09 (1.14)	-0.48 (-1.62)	2.07*** (3.31)
Non-ATMF caps	4wks	0.01 (0.12)	0.00 (0.04)	0.04 (1.01)	0.06 (0.52)	0.11 (0.60)
	8wks	-0.08 (-0.43)	-0.01 (-0.09)	0.07 (0.87)	0.08 (0.35)	0.32 (0.77)
	16wks	-0.45 (-1.22)	-0.18 (-0.89)	0.13 (0.92)	-0.11 (-0.29)	0.62 (0.70)

Notes: Pairwise comparisons of the models' forecasting performance 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, "3C" denotes the $N=3$ cap model, and "RW" denotes the random walk model for interest rates and derivatives prices. Interest rate forecast errors are measured in basis points while derivative forecast errors are measured in percentages.

Table 10: Statistical comparisons of forecasting performance

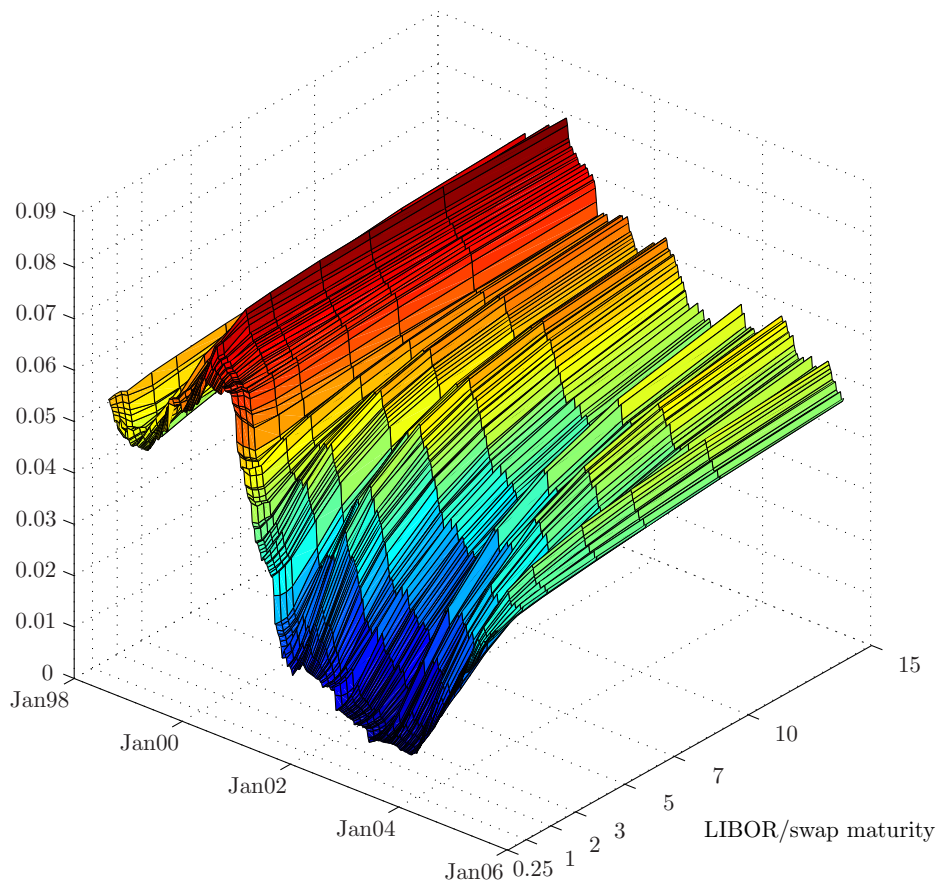


Figure 1: Time series of LIBOR and swap rates

Each time series consists of 360 weekly observations from August 21, 1998 to July 8, 2005. Source: Bloomberg.

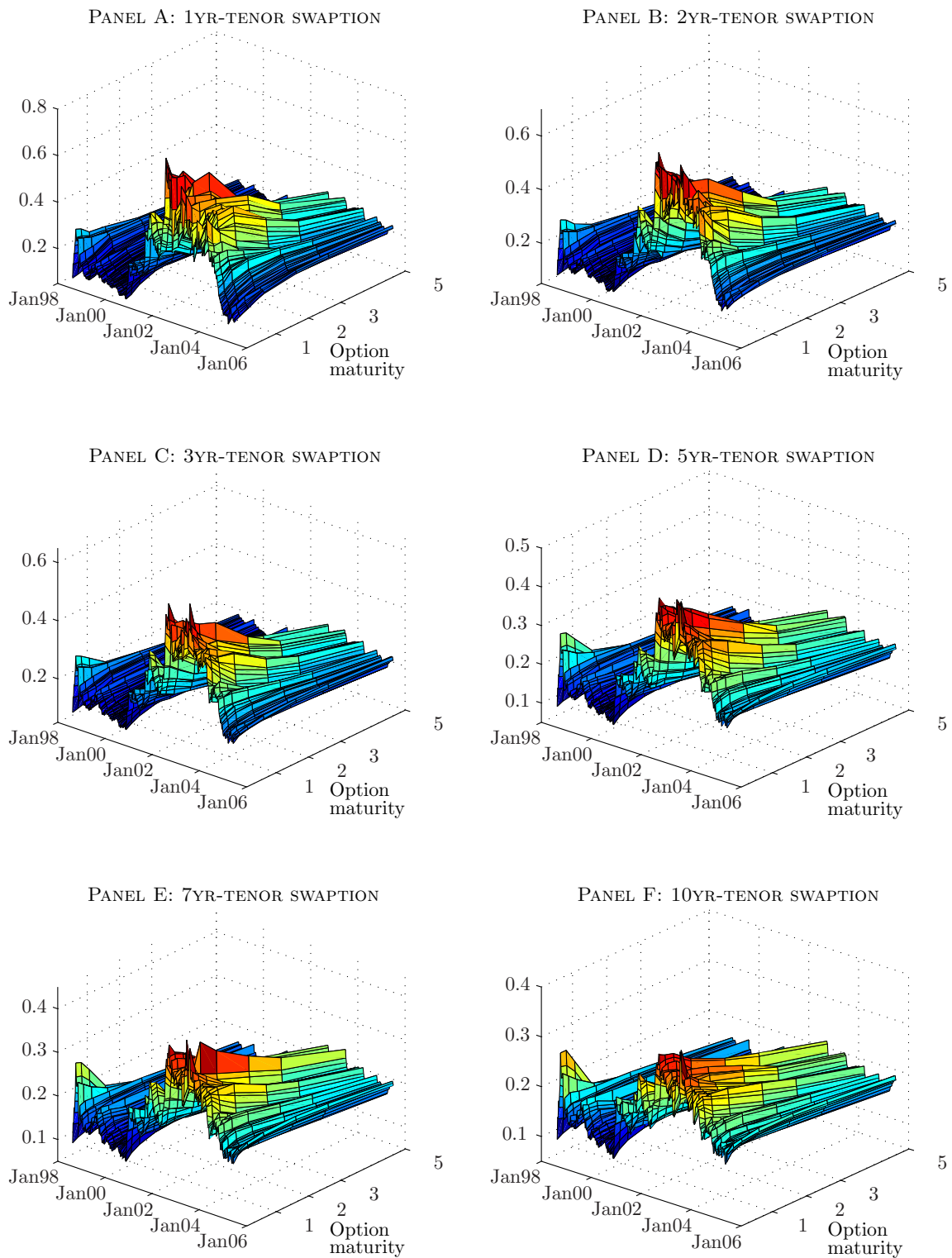


Figure 2: Time series of log-normal implied ATMF swaption volatilities
 Each time series consists of 360 weekly observations from August 21, 1998 to July 8, 2005. Source: Bloomberg.

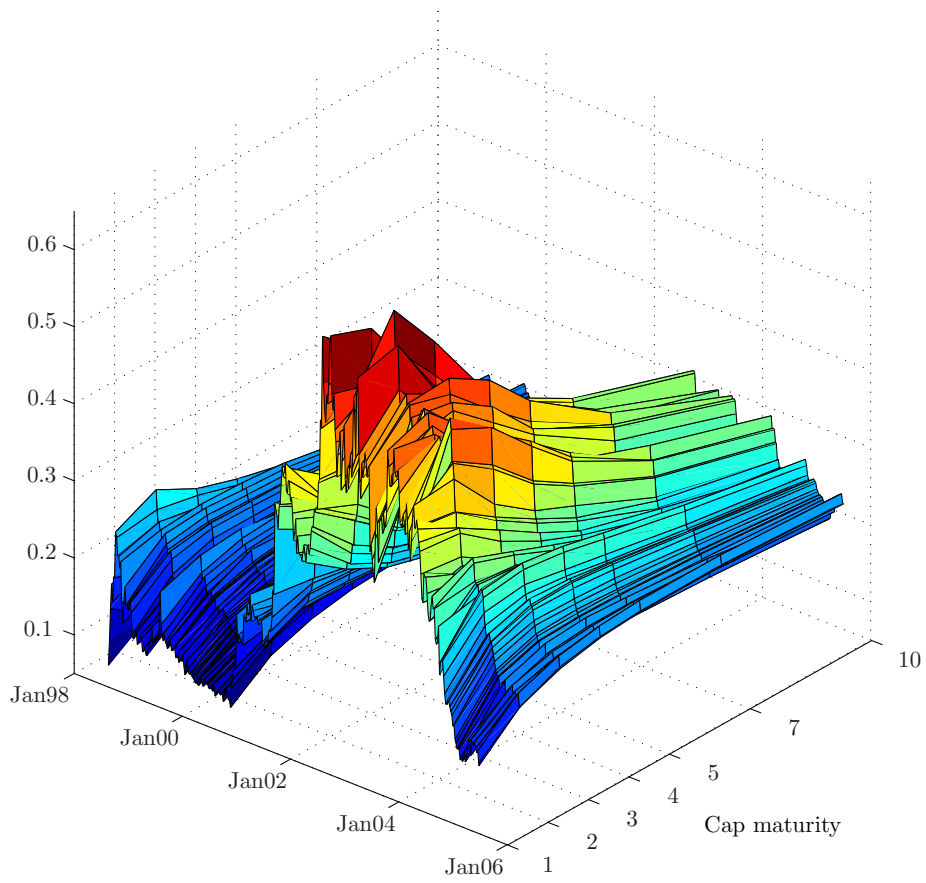


Figure 3: Time series of log-normal implied ATM cap volatilities

Each time series consists of 360 weekly observations from August 21, 1998 to July 8, 2005. Source: Bloomberg.

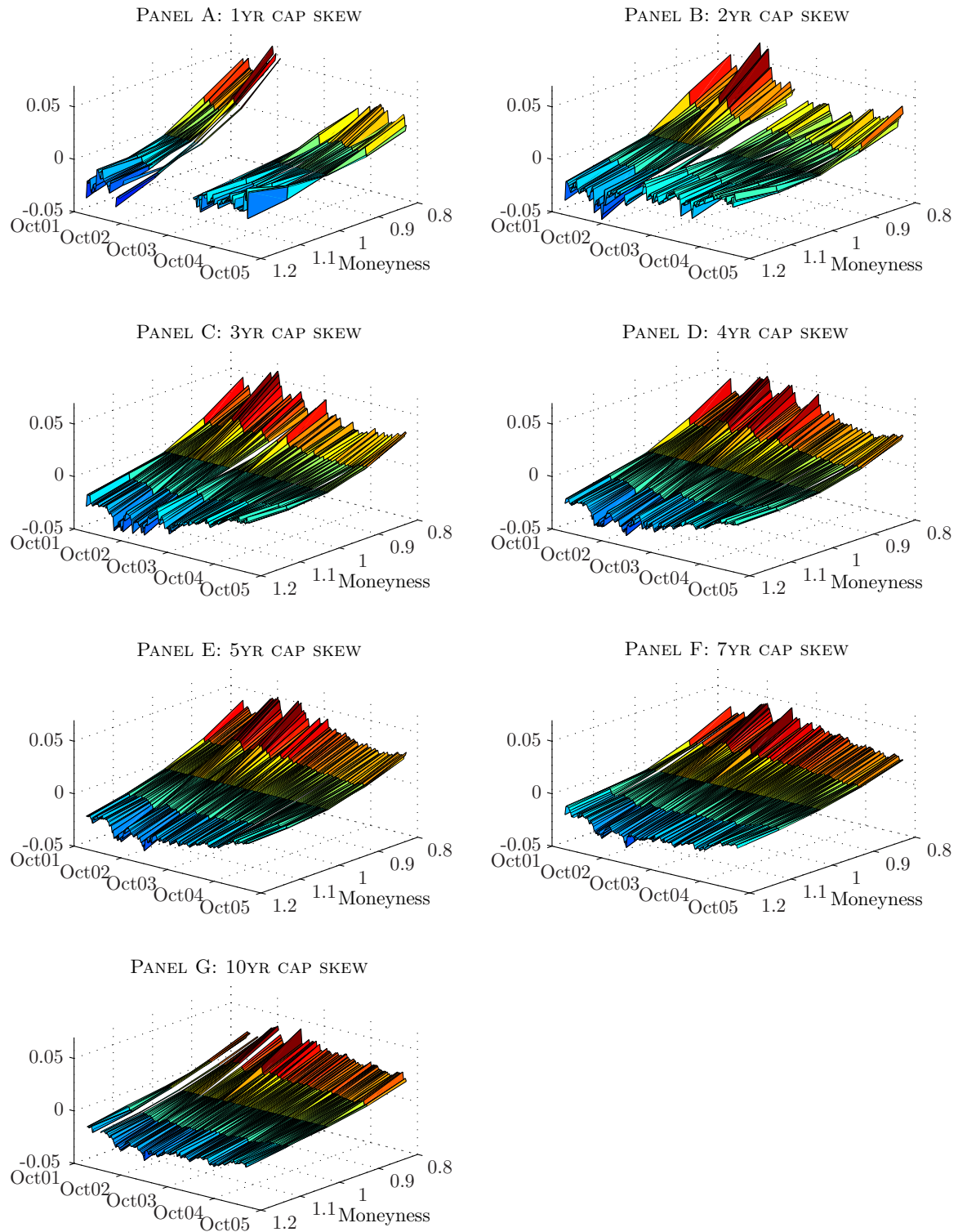


Figure 4: Time series of log-normal implied cap skews

The skews are the differences between the implied volatilities across moneyness and the implied volatilities of the corresponding ATM caps. “Moneyness” of a given cap is defined as the ratio between its strike and the strike on the ATM cap with the same maturity. Each time series consists of a maximum of 184 weekly observations from January 4, 2002 to July 8, 2005. Source: Bloomberg.

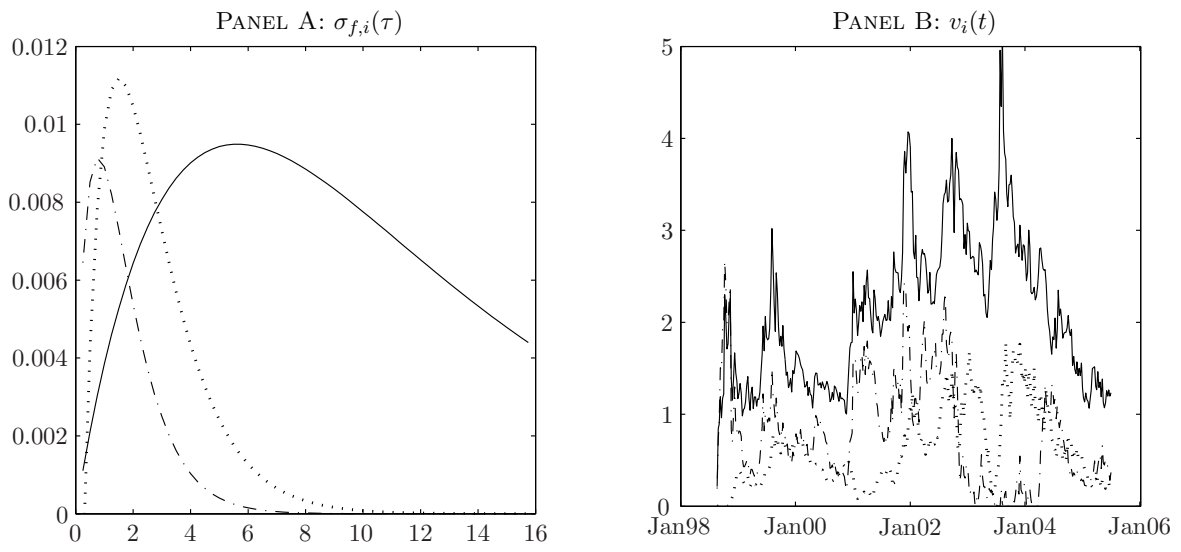


Figure 5: $\sigma_{f,i}(\tau)$ and $v_i(t)$ for the $N=3$ swaption and cap model

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

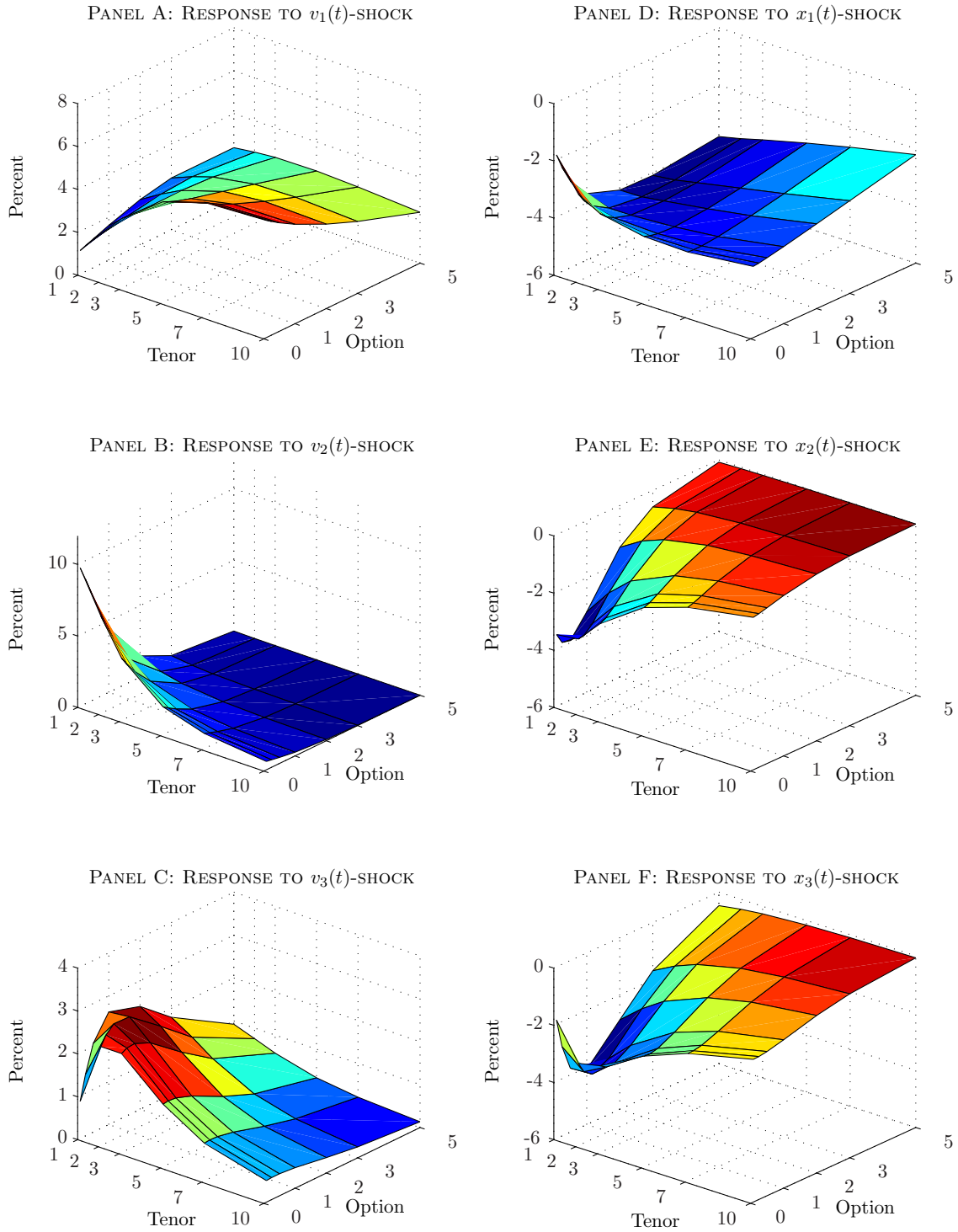
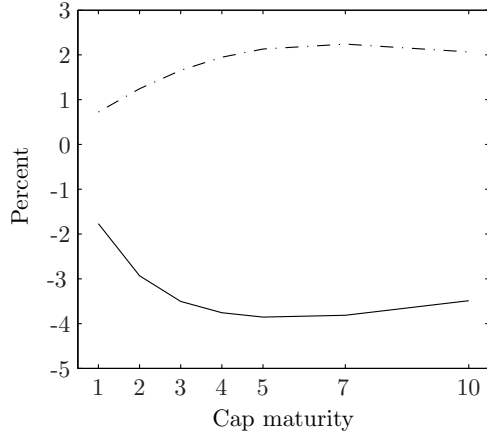


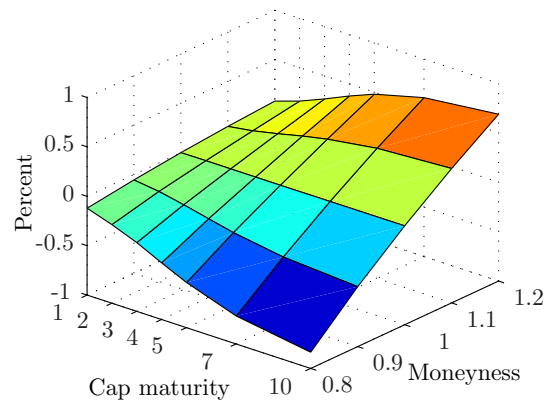
Figure 6: Dynamics of the log-normal implied swaption volatility surface

Panel A, B and C show the impact on the log-normal implied swaption volatility surface from one-standard deviation shocks to $v_1(t)$, $v_2(t)$ and $v_3(t)$, respectively. Panel D, E and F show the impact on the log-normal implied swaption volatility surface from one-standard deviation shocks to $x_1(t)$, $x_2(t)$ and $x_3(t)$, 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 caps model.

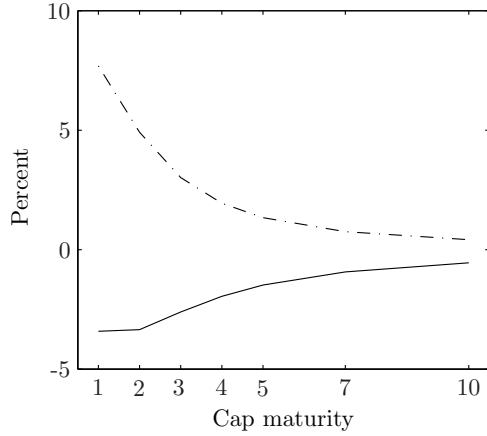
PANEL A: RESPONSE TO $x_1(t)$ AND $v_1(t)$ SHOCKS



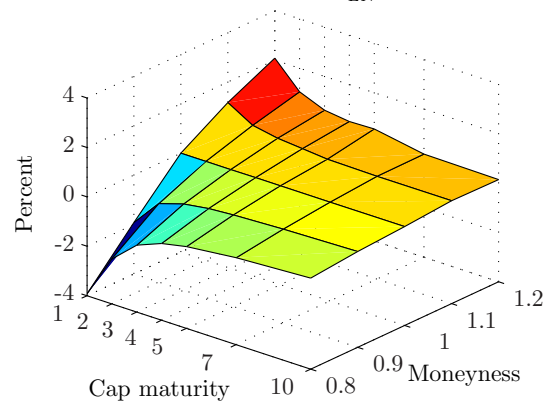
PANEL D: $\partial(\sigma_{LN} - \sigma_{LN}^{ATMF})/\partial\rho_1$



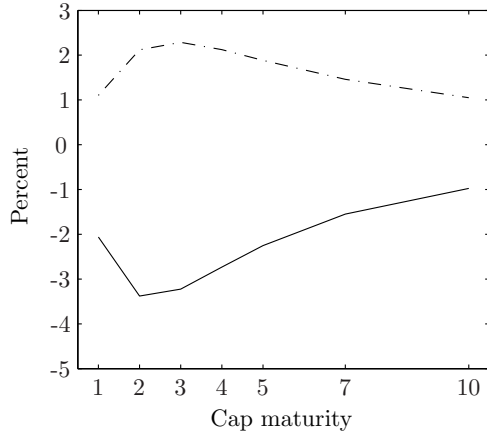
PANEL B: RESPONSE TO $x_2(t)$ AND $v_2(t)$ SHOCKS



PANEL E: $\partial(\sigma_{LN} - \sigma_{LN}^{ATMF})/\partial\rho_2$



PANEL C: RESPONSE TO $x_3(t)$ AND $v_3(t)$ SHOCKS



PANEL F: $\partial(\sigma_{LN} - \sigma_{LN}^{ATMF})/\partial\rho_3$

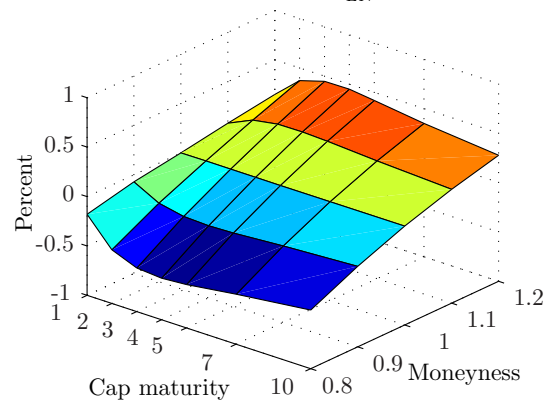


Figure 7: Dynamics of the log-normal implied cap skew volatility surface

Panel A, B and C show the impact on log-normal implied ATMF cap volatilities from one-standard deviation shocks to $x_1(t)$ and $v_1(t)$, $x_2(t)$ and $v_2(t)$, and $x_3(t)$ and $v_3(t)$, respectively. '—' denotes shocks to $x_i(t)$ and '- · -' denotes shocks to $v_i(t)$. Panel D, E and F show the derivatives of the differences between non-ATMF and ATMF log-normal 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 caps model.

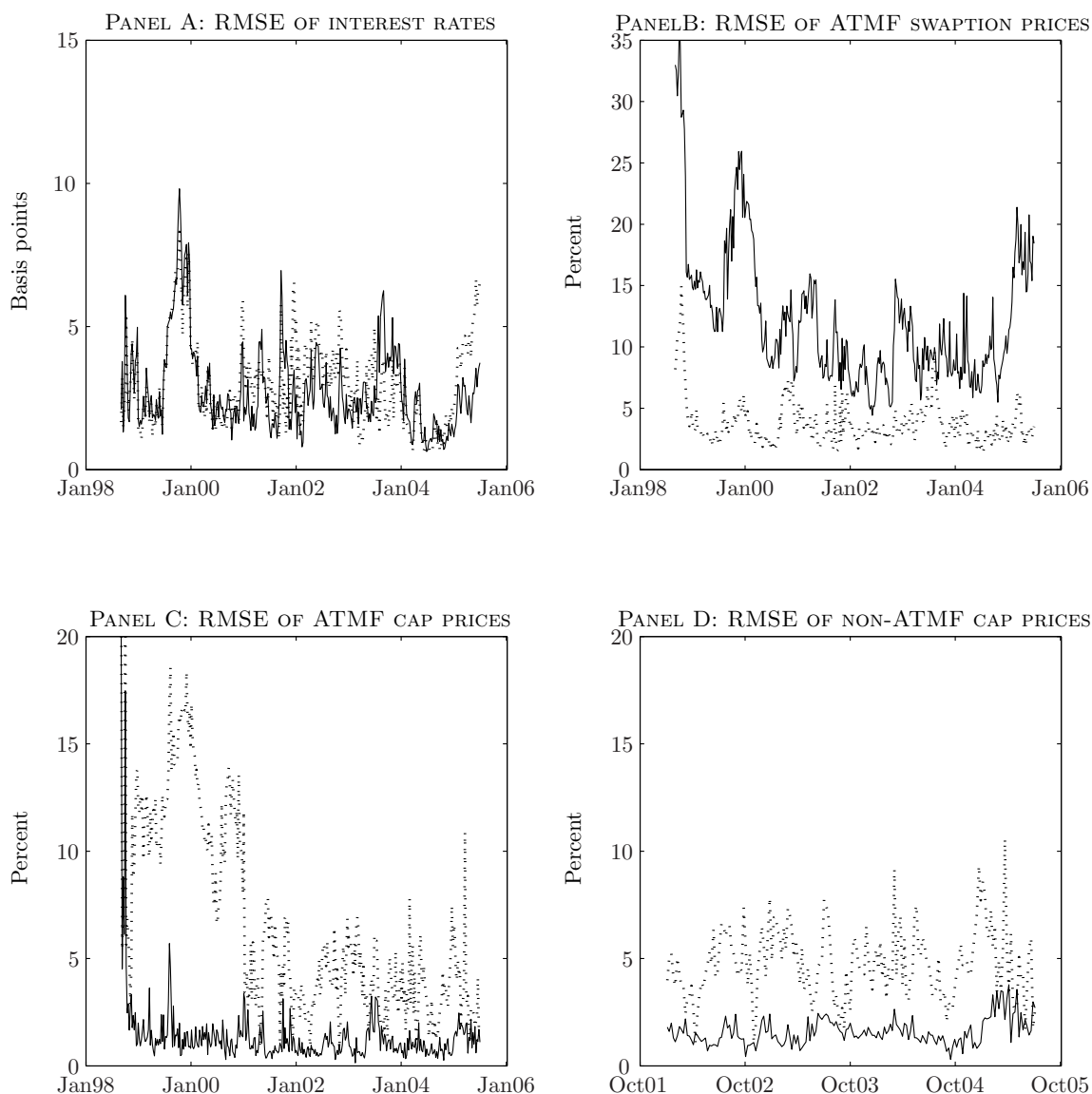


Figure 8: Time-series of RMSEs for interest rates, swaptions and caps

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 panel 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.

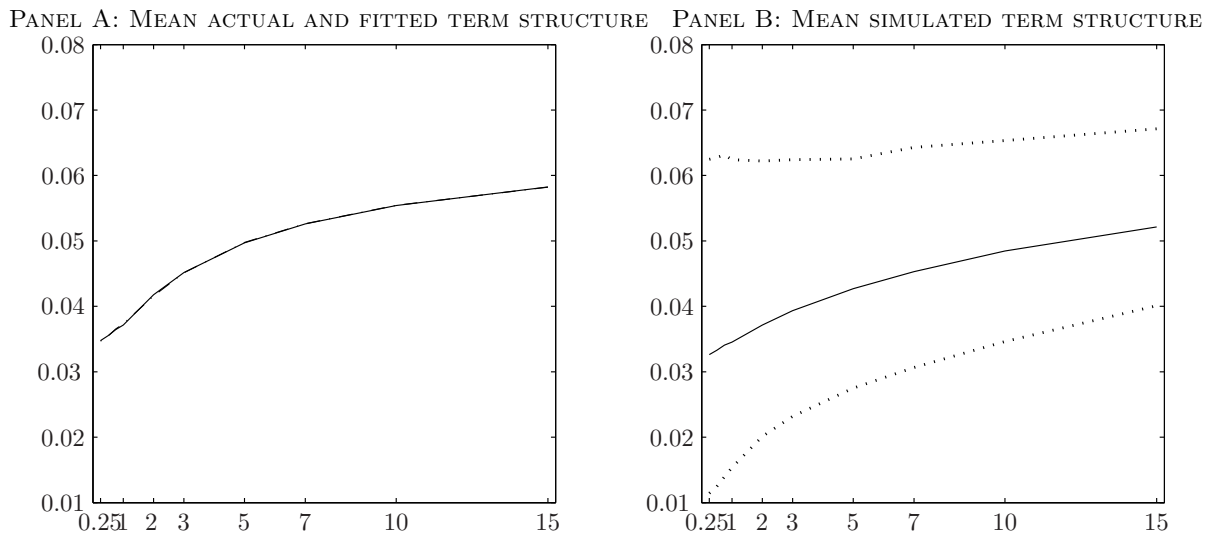


Figure 9: Means of actual, fitted and simulated term structures

Panel A shows the means of the actual (‘—’) and fitted (‘- · -’) term structures in the case of the $N=3$ swaptions and caps model. Means are computed over 360 weekly observations from August 21, 1998 to July 8, 2005. In Panel B we first simulate 1000 term structure samples each of length of 360. We then compute the mean term structure for each sample to obtain the small-sample distribution of the mean term structure generated by the model. Panel B shows the mean (‘—’) and the 2.5 and 97.5 percentiles (‘- · -’) of this distribution.

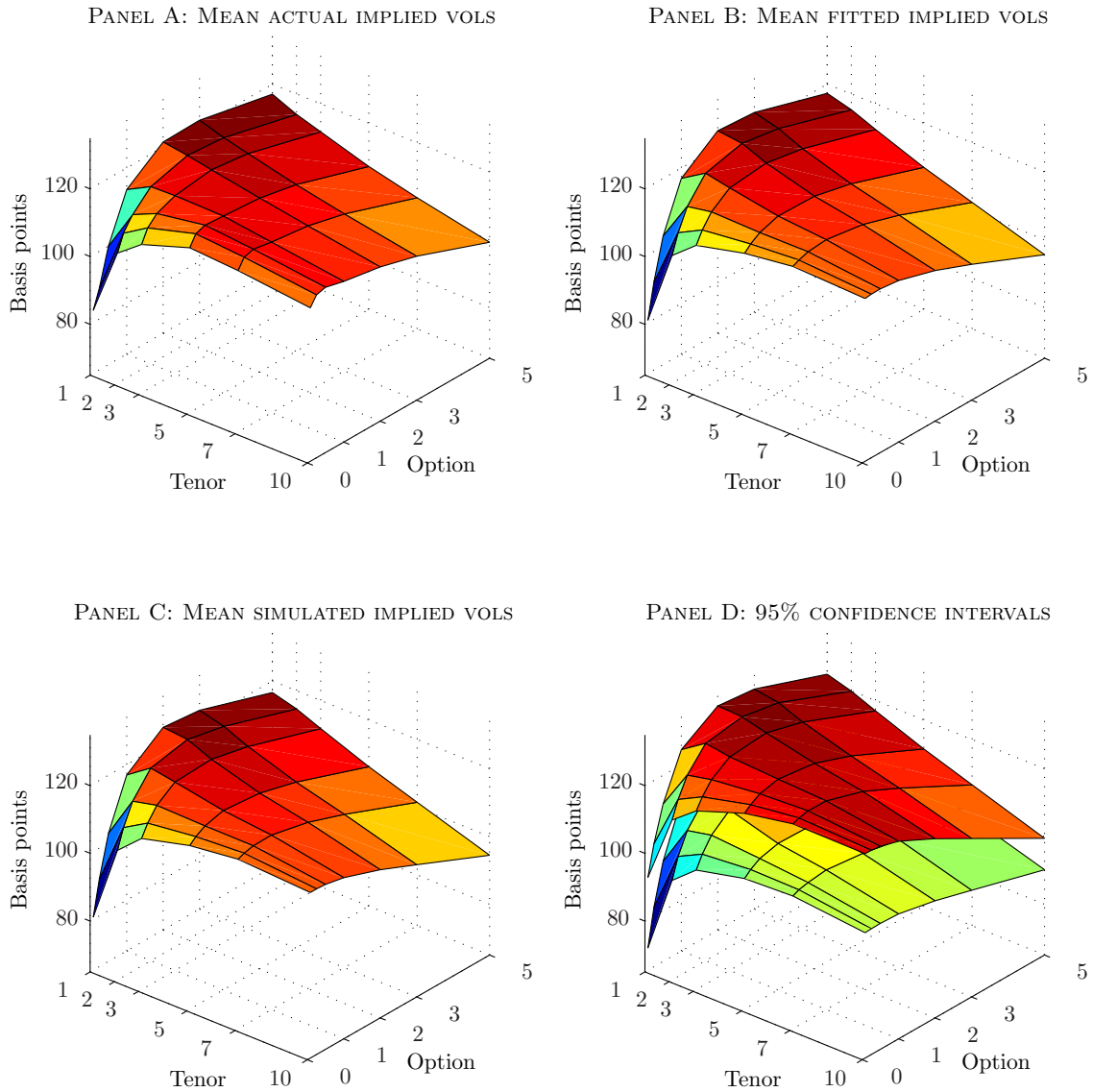


Figure 10: Means of actual, fitted and simulated normal implied swaption volatility surfaces. Panel A shows the mean of the actual normal implied volatility surface. Panel B shows the mean of the fitted normal implied volatility surface in the case of the $N=3$ swaptions and caps model. Means are computed over 360 weekly observations from August 21, 1998 to July 8, 2005. In Panel 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 2.5 and 97.5 percentiles of this distribution.

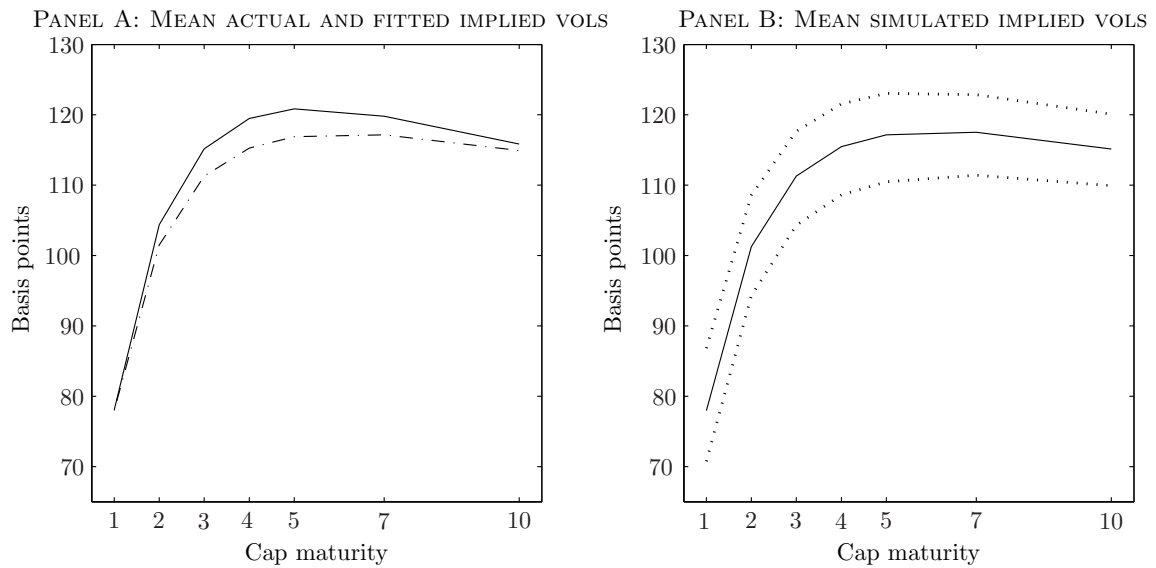


Figure 11: Means of actual, fitted and simulated normal implied cap volatilities

Panel A shows the means of the actual ('—') and fitted ('- · -') normal implied ATMF cap volatilities in the case of the $N=3$ swaptions and caps model. Means are computed over 360 weekly observations from August 21, 1998 to July 8, 2005. In Panel B we first simulate 1000 samples, each of length of 360, of normal implied ATMF cap volatilities. We then compute the mean volatilities for each sample to obtain the small-sample distribution of the mean volatilities generated by the model. Panel B shows the mean ('—') and the 2.5 and 97.5 percentiles ('...') of this distribution.

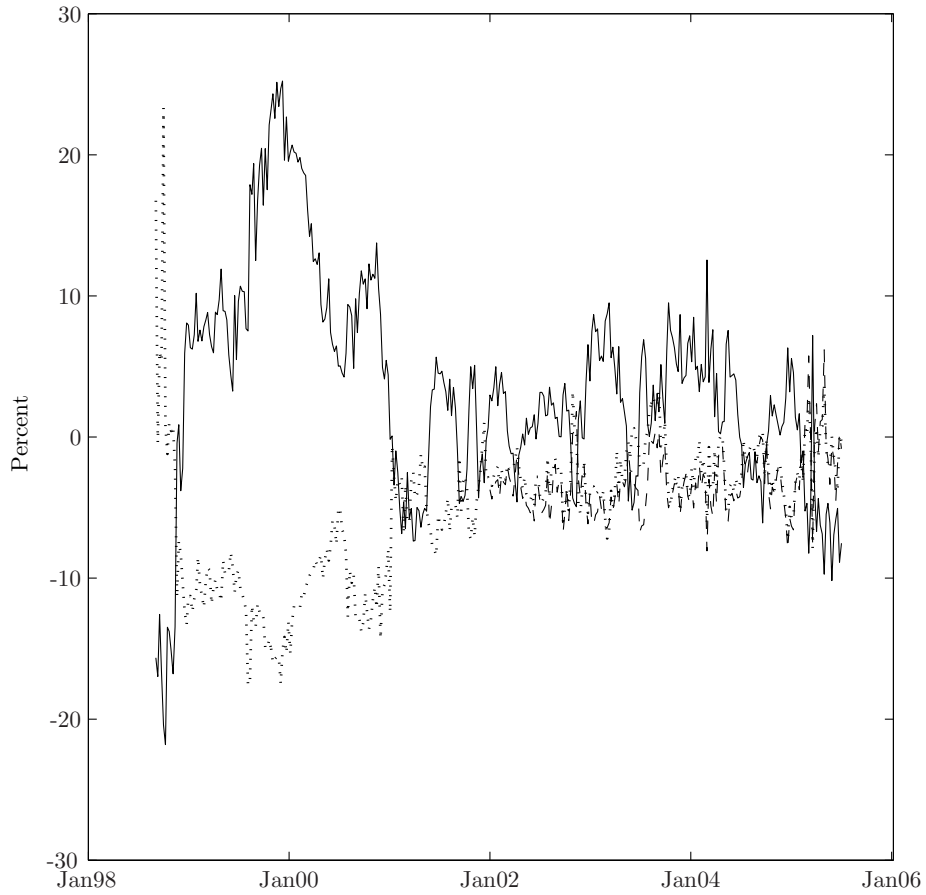


Figure 12: Time-series of mis-valuations of caps and swaptions

'—' 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 ten 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.

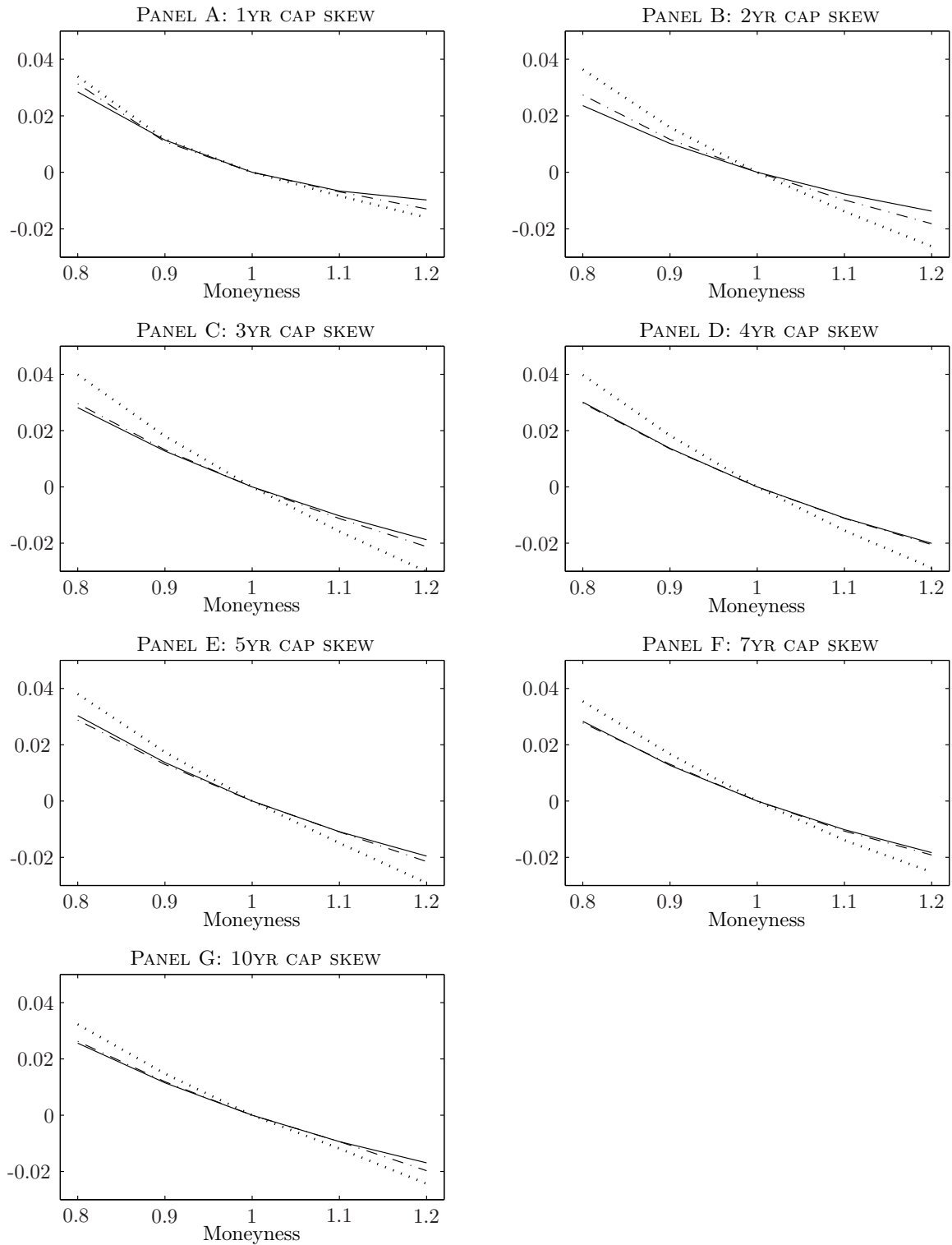


Figure 13: Average fit to log-normal implied cap skews

'—' denotes the average of the actual skews. '- - -' denotes the average of the fitted skews for the $N=3$ model estimated on caps data. '...' denotes the average of the fitted skews for the $N=3$ model estimated on caps 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 ATM caps. "Moneyness" of a given cap is defined as the ratio between its strike and the strike on the ATM 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.

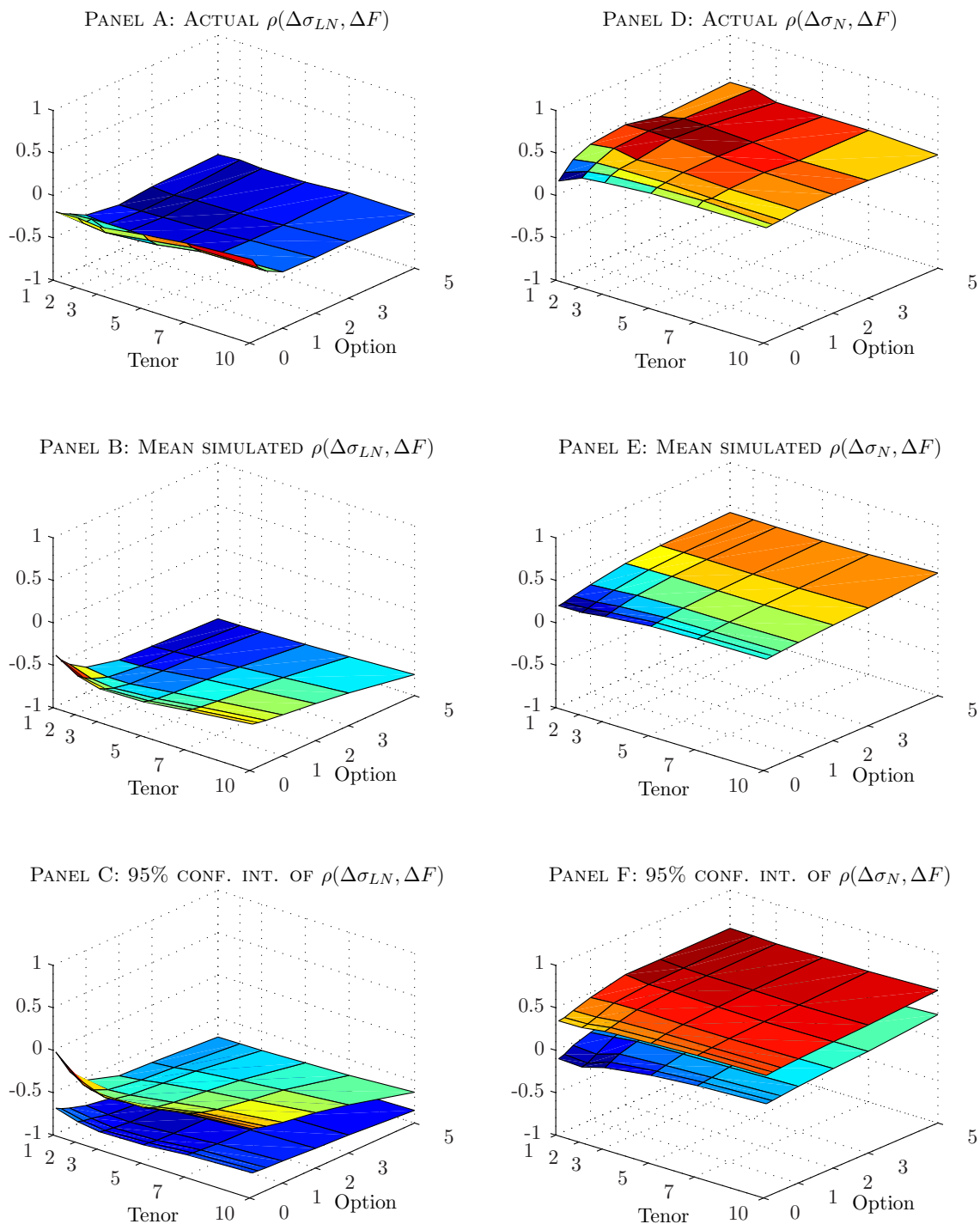


Figure 14: Reproducing the implied volatility – interest rate correlations

Panel A shows the actual correlations between changes in log-normal 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 Panel B, C, E and F we first simulate 1000 samples, each of length of 360, of log-normal and normal implied swaption volatilities and the underlying forward swap rates. 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 parameters generated by the model. Panel B and C show the mean and 2.5 and 97.5 percentiles, respectively, of the $\rho(\Delta\sigma_{LN}, \Delta F)$ -distributions while Panel E and F show the mean and 2.5 and 97.5 percentiles, respectively, of the $\rho(\Delta\sigma_N, \Delta F)$ -distributions.

References

- ANDERSEN, L. AND R. BROTHERTON-RATCIFFE (2005): “Extended Libor market models with stochastic volatility,” *Journal of Computational Finance*, 9.
- 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.
- 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*, 3:167–179.
- BLYTH, S. AND J. UGLUM (1999): “Rates of skew,” *Risk*.
- BRACE, A., D. GATAREK, AND M. MUSIELA (1997): “The market model of interest rate dynamics,” *Mathematical Finance*, 7:127–155.
- CASASSUS, J., P. COLLIN-DUFRESNE, AND R. GOLDSTEIN (2005): “Unspanned stochastic volatility and fixed income derivatives pricing,” *Journal of Banking and Finance*, 29:2723–2749.
- CHEREDITO, P., D. FILIPOVIC, AND R. KIMMEL (2003): “Market price of risk specifications for affine models: Theory and evidence,” Working paper, Princeton University, forthcoming in *Journal of Financial Economics*.
- 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*, 5:129–155.
- COLLIN-DUFRESNE, P. AND R. GOLDSTEIN (2002a): “Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility,” *Journal of Finance*, 57:1685–1730.
- (2002b): “Pricing swaption within an affine framework,” *Journal of Derivatives*, pages 1–18.
- (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*, 55:1943–1978.
- (2002): “Expectations puzzles, time-varying risk premia, and affine models of the term structure,” *Journal of Financial Economics*, 63:415–442.
- 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*, 34:131–157.
- DIEBOLD, F. AND R. MARIANO (1995): “Comparing predictive accuracy,” *Journal of Business and Economic Statistics*, 13:253–263.
- DUFFEE, G. (2002): “Term premia and interest rate forecasts in affine models,” *Journal of Finance*, 57:405–443.
- 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*, 68:1343–1376.
- 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 University Press, Princeton, New Jersey.
- HAN, B. (2004): “Stochastic volatilities and correlations of bond yields,” Working paper, Ohio State University.
- 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*, 60:77–105.
- 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*, 6:327–343.
- HO, T. AND S. LEE (1986): “Term structure movements and pricing interest rate contingent claims,” *Journal of Finance*, 41:1011–1029.
- HULL, J. AND A. WHITE (1990): “Pricing interest-rate-derivative securities,” *Review of Financial Studies*, 3:573–592.
- JAMSHIDIAN, F. (1997): “Libor and swap market models and measures,” *Finance and Stochastics*, 1:293–330.
- JARROW, R., H. LI, AND F. ZHAO (2004): “Interest rate caps ‘smile’ too! But can the LIBOR market models capture it?,” Working paper, Cornell University, forthcoming in *Journal of Finance*.
- LI, H. AND F. ZHAO (2006): “Unspanned stochastic volatility: Evidence from hedging interest rate derivatives,” *Journal of Finance*, 61:341–378.
- 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.
- LUND, J. (1997): “Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates,” Working paper, Aarhus School of Business.
- MILTERSEN, K., K. SANDMANN, AND D. SONDERMANN (1997): “Closed form solutions for term structure derivatives with log-normal interest rates,” *Journal of Finance*, 52:409–430.

- MUNK, C. (1999): “Stochastic duration and fast coupon bond option pricing in multi-factor models,” *Review of Derivatives Research*, 3:157–181.
- NELSON, C. AND A. SIEGEL (1987): “Parsimonious modeling of yield curves,” *Journal of Business*, 60:473–489.
- NEWKEY, W. AND K. WEST (1987): “A simple, positive semi-definit, heteroscedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55:703–708.
- RITCHKIN, P. AND I. CHUANG (1999): “Interest rate option pricing with volatility humps,” *Review of Derivatives Research*, 3:237–262.
- SCHRAGER, D. AND A. PELSSER (2005): “Pricing swaption in affine term structure models,” Working paper, Erasmus University, forthcoming in *Mathematical Finance*.
- SINGLETON, K. AND L. UMANTSEV (2002): “Pricing coupon-bond options and swaptions in affine term structure models,” *Mathematical Finance*, 12:427–446.
- 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*, 17:131–160.
- 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.
- ZHOU, F. (2003): “Black smirks,” *Risk*.