

A Swaption Volatility Model using Markov Regime Switching

Richard White* and Riccardo Rebonato*^{†‡}

* The Royal Bank of Scotland, 135 Bishopsgate, London EC2 3UR

† Oxford University (OCIAM)

‡ Imperial College, Tanaka Business School, London

February 16, 2007

Abstract

We present a parsimonious, financially-motivated model which provided a good description of the swaption volatility matrix. The core model consists of a hidden Markov chain with two volatility states, *normal* and *excited*. Each state has its own curve for the instantaneous forward volatility. The volatilities in the swaption matrix result from averaging over all possible paths along the Markov chain. We provide a fast, accurate, analytic method for calculating the swaption matrix from this model. With this procedure we show dramatic improvements over the Rebonato approach[1] in the quality of fits when the market appears to be excited.

1 Introduction

The 2004 paper by Rebonato[1], *Forward-rate Volatilities and the Swaption Matrix*, assumes that the instantaneous volatility of the forward rates is a deterministic function of time to maturity only. The model was fitted to USD at-the-money swaption volatility market data, and despite the restrictive assumption, the recovery of market prices was surprisingly good during *normal* market conditions. As discussed below, in this work we extend this study to EUR swaptions, and we find that the agreement is even better. However, the original work on USD data also highlighted clear systematic shortcomings during periods when the market is *excited*. The point was made that these shortcomings are unavoidable for any underlying if the relevant volatility functions are assumed to be functions of time or of time to expiry only¹.

In this paper we apply this model to a longer period of

¹Indeed, Rebonato's paper was subtitled: *Why Neither Time Dependence nor Time Homogeneity Will Do*.

USD at-the-money swaption volatility market data and observe the same shortcomings seen in [1]. In Section 6.1 we show that the deterministic model describes the Euro market extremely well. The data in that case is only from January 2001 so does not include the market turbulence of 1998.

The problem with this deterministic volatility model is illustrated by Figures 1 and 2. Figure 1 on 21-Jul-98 shows a *normal* market situation, and it is apparent that a small number of parameters (five, in this particular case) do a good job of recovering the volatility matrix in its entirety. Four months later, the market is *excited*, and the fit is shown in Figure 2. Here the volatility at the short end (6 month into 1 year) is $\sim 22\%$ while the long end (10 years into 1 year) is 10%. The deterministic time-homogeneous volatility model attempts to fit this market by having an instantaneous volatility of the forward rates that is much higher close to maturity than far from it, however this results in significant mispricing of options in common expiry blocks. One could, of course, improve the fits to the short-dated options, but this would systematically misprice the longer-dated ones, and vice versa. Clearly the situation is more complex than allowed for by the model. One can interpret the mispricing pattern as indicating that the market recognises the onset of an *excited* period, but 'believes' that this period will subside and revert to *normal* conditions in a relatively short time. This is reflected in higher volatilities for short-dated options, but barely changed volatilities for longer-dated ones. If this interpretation is correct, clearly no set of time-homogeneous parameters can describe this state of affairs: a calibration that accounted for the *excited* short-dated option prices would overestimate the volatilities of long-dated options because it cannot 'know' about reversion to a *normal* state; and vice versa. This is the essential feature that we want to

capture with the model proposed in this work.

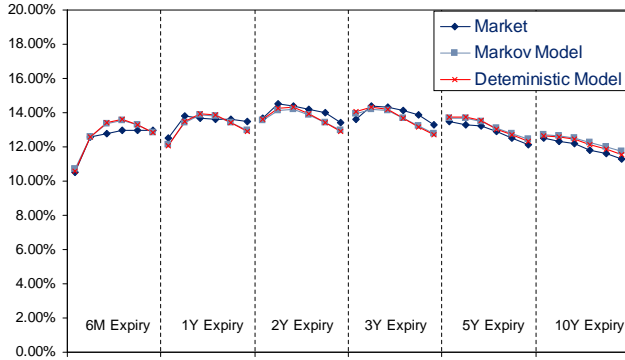


Figure 1: The USD swaption matrix for 21-July-1998, fitted with the original deterministic vol model[1] and the new Markov model presented in this paper. Each expiry block shows the implied volatility for the options into 1-year, 2-year, 3-year, 5-year, 7-year and 10-year tenors. Here the market is in a *normal* state and the deterministic model fits reasonably well, so there is little gained from the added complexity of the Markov model.

In order to give a more satisfactory description of the swaption market under all market conditions, we generalise the original deterministic volatility model as follows. The basic assumption is that the market can be in either a *normal* or an *excited* state; in each state there exists a deterministic function of the residual time to maturity for the instantaneous volatility of the forward rates. The world switches between these two states according to a hidden Markov chain, so that the resulting swaption prices are the result of averaging over all possible paths along the Markov chain.

Both models have been fitted to 8.5 years of USD swaption data (with a 6x6 swaption matrix), and close agreement between the Markov model and the market is found across the wide range of market conditions that occurred during this period.

The rest of the paper is organised as follows: The following sections outline the building blocks of the deterministic volatility model[1]. Then sections 2.3 to 2.4 show how to extend this model with a two-state hidden Markov chain governing with of two volatility curves drive the forwards rates over a given period. In section 3 we derive a fast, reliable method for calculating the swaption prices under the Markov chain model, and in

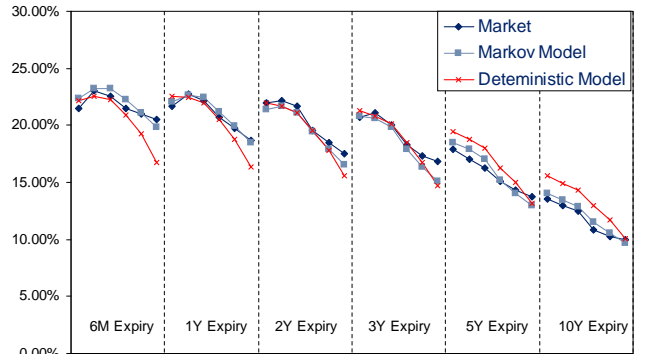


Figure 2: The USD Swaption matrix for 18-Nov-1998, fitted with the original deterministic vol model[1] and the new Markov model presented in this paper. The market here is in an *excited* state, but believes we will revert to the *normal* state. In this case, the Markov model presented in this paper is a much better fit than the deterministic model.

section 4 we demonstrate the accuracy of this method. Section 5 details the data sets and how the models are calibrated to this data. Finally section 6 discusses the results for both the Euro and USD data sets, and section 7 gives conclusions.

2 The Models

In the original model there is one volatility curve that is common to all forward rates - different forward rates experience different instantaneous volatilities only depending on their own time to maturity. As discusses in the introduction, this single curve is inadequate during periods of market excitation.

A natural extension is to have two volatility curves, with the ‘world’ flipping between them according to a Hidden-Markov-chain. This extension gives a much richer set of outputs (Swaption Matrices) at the expense of abandoning the closed-form approximate solution one has with the single curve [2].

2.1 Forward Rates

In this model, the one-period forward rates $F(t, T_i, T_{i+1}) \equiv F_i(t)$ are assumed to follow dis-

placed diffusion processes²[7] in their terminal measures \mathbb{T}_i [10]. It has been shown that displaced diffusion can be used as a computationally simple proxy for a CEV process [9, 8]. The dynamics of $F_i(t)$ in the terminal measures \mathbb{T}_i are given by:

$$\frac{dF_i(t)}{F_i(t) + \epsilon} = \sigma_i(t, T_i) dW_i^{\mathbb{T}_i} \quad (1)$$

with $\sigma_i(t, T_i)$ a deterministic instantaneous volatility that is a function of time to maturity only as follows

$$\sigma_i(t, T_i) = (a + b\tau)e^{-c\tau} + d \quad \text{with } \tau = T_i - t \quad (2)$$

This is a widely used functional form[1, 3]. With c positive, the volatility tends to d for large time-to-maturity τ , and is equal to $a + d$ at $\tau = 0$ which corresponds to option expiry. When $b, c > 0$ there is a maximum value at

$$\tilde{\tau} = 1/c - a/b \quad (3)$$

so since $\sigma_i(t, T_i)$ is only relevant when $\tau \geq 0$, $\sigma_i(t, T_i)$ is a monotonically decreasing function of τ when $b < ac$ provided $a > 0$, and is humped provided $b > \max(0, ac)$. Examples of this function are shown in figure 3. The rationale for a ‘humped’ volatility curve are given in [1, 4]. In practice, when fitted to calm periods in the data, the maximum of the curve is typically found between 1.5 and 3 years to maturity (see figure 19). The types of volatility curves shapes obtained from calibration to data are discussed in the results section starting on page 10.

The displacement ϵ and the parameters a, b, c, d of the volatility curve are common to all forward rates. In addition, the stochastic drivers for each forward rate i and j have a correlation ρ_{ij} that we take as one parameter function of the difference in maturities only:

$$\rho_{ij} = e^{-\lambda|T_i - T_j|} \quad \text{with } \lambda \geq 0 \quad (4)$$

so that there is a de-correlation as the difference in maturities increases.

2.2 Swaption Volatilities

The swap rate for a swap starting at time T_α and ending at T_β can be written as a weighted sum of the underlying

²In displaced-diffusion, the quantity $F + \epsilon$ follows a log-normal process, where ϵ is some fixed quantity with $\epsilon \geq 0$. When the displacement $\epsilon = 0$ we have a log-normal process and when $\epsilon \rightarrow \infty$ we have a normal process. See appendix A for a brief discussion of displaced diffusion.

one-period forward rates³ $F_i(t)$:

$$SR_{\alpha\beta}(t) = \sum_{i=\alpha}^{\beta-1} w_i^{\alpha\beta}(t) F_i(t) \quad (5)$$

$$\text{where } w_i^{\alpha\beta}(t) = P_{i+1}(t)/X^{\alpha\beta}(t) \quad (6)$$

$$\text{and } X^{\alpha\beta}(t) = \sum_{i=\alpha+1}^{\beta} P_i(t) \quad (7)$$

Here $P_i(t) \equiv P(t, T_i)$ is the discount factor for time T_i seen at time t . This form is useful because the weights $w_i^{\alpha\beta}(t)$ only have a weak dependence on the forward rates.

If we assume that the swap rate follows a displaced diffusion process with displacement ϵ' , $dSR_t = [SR_t + \epsilon']\sigma_t^{\epsilon'} dZ_t$, then the instantaneous volatility can be approximated by⁴[4, 2]:

$$[\sigma_t^{\epsilon'}]^2 \approx \sum_{i,j=\alpha}^{\beta-1} \zeta_i^{\alpha\beta}(t) \zeta_j^{\alpha\beta}(t) [F_i(t) + \epsilon][F_j(t) + \epsilon] \rho_{ij}(t) \sigma_i(t) \sigma_j(t) \quad (8)$$

where the double sum is over all the forwards active in the swap, $\rho_{ij}(t)$ is the instantaneous correlation between the driver of the i^{th} and j^{th} one-period forwards (i.e. $dW_i(t)dW_j(t)/dt$) and $\zeta_i^{\alpha\beta}(t)$ is:

$$\zeta_i^{\alpha\beta}(t) = \frac{w_i^{\alpha\beta}(t) + \sum_{j=\alpha}^{\beta-1} F_j(t) \frac{\partial w_j^{\alpha\beta}}{\partial F_i}}{SR_{\alpha\beta}(t) + \epsilon'} \quad (9)$$

It is a good approximation to assume $\zeta_i^{\alpha\beta}(t) = \zeta_i^{\alpha\beta}(0) \forall t$ [2]. Also the second term in the numerator is small and can be ignored - the term is exactly zero if the yield curve is flat.

Defining $\nu^{\epsilon'}$ as the displaced diffusion variance then

$$\nu^{\epsilon'}(a, b, c, d, \epsilon) = \int_0^{T_\alpha} [\sigma_t^{\epsilon'}]^2 dt = \sum_{i,j=\alpha}^{\beta-1} \left(\zeta_i^{\alpha\beta}(0) \zeta_j^{\alpha\beta}(0) [F_i(0) + \epsilon][F_j(0) + \epsilon] \times \int_0^{T_\alpha} \rho_{ij}(t) \sigma_i(t) \sigma_j(t) dt \right) \quad (10)$$

³For this swap the LIBOR rates are set at times $T_\alpha, T_{\alpha+1}, \dots, T_{\beta-2}, T_{\beta-1}$ and payments are made at times $T_\alpha + 1, T_{\alpha+2}, \dots, T_{\beta-1}, T_\beta$

⁴The case that both the forward rates and the swap rate are log-normal cannot be exactly true. However the approximate is very close. See [12, Need to check this reference] for a full discussion.

where we have emphasised the parameter dependence. Appendix A gives the price of an option under displaced diffusion. This can then be used to find the swaption Black-vol $\sigma_{\text{Black}}^{\alpha\beta}$ by inverting the Black formula. For at-the-money options the relation $\sigma_{\text{Black}} \approx (F + \epsilon)\sigma^\epsilon/F$ holds and the dependence on ϵ' is cancelled out. So one can finally write:

$$\left[\sigma_{\text{Black}}^{\alpha\beta}(a, b, c, d, \epsilon)\right]^2 T_\alpha \approx \sum_{i,j=\alpha}^{\beta-1} \left(\zeta_i^{\alpha\beta}(0)\zeta_j^{\alpha\beta}(0)[F_i(0) + \epsilon][F_j(0) + \epsilon] \times e^{-\lambda|T_i - T_j|} \int_0^{T_\alpha} \sigma_i(t)\sigma_j(t)dt \right) \quad (11)$$

where the ϵ' of equation 9 is set to zero.

Clearly the integral $\int_0^T \sigma_i(t)\sigma_j(t)dt \equiv \nu_{ij}(T; a, b, c, d)$ is of critical importance. For the form of the instantaneous volatility given in equation 2, $\nu_{ij}(T)$ has an analytic, if cumbersome form (see page 172 of Rebonato[4]).

With this model, the entire swaption matrix is described by just 6 parameters a, b, c, d, ϵ and λ . However when fitting USD market data, one usually finds that $\lambda \approx 0$ (i.e. $\rho_{ij} \approx 1$ for all forward pairs), and that the displacement $\epsilon \gg 1$, which implies a normal model. If λ and ϵ are fixed at 0 and 1 respectively, then we effectively only have four parameters. This is not the case with the Euro market, where all 6 parameters are relevant.

2.3 Hidden Markov Chain Model

As discussed in the introduction, the original deterministic volatility model does not describe the market well in times when the market is in a temporary *excited* state. To extend the model we propose that two volatility curves exist: an *excited* curve and a *normal* curve. At any given time the market is in one of these states, and transitions between these states are governed by a hidden Markov chain.

Markov chain models have been used already to describe option prices (e.g. [11]). The task we are facing is however more challenging, because in the interest rate context we have to deal with transitions between *functions* rather than level. The numerical approximations we devise are therefore totally different. In our model the parameters describing the instantaneous volatility can take two sets of values $(a_n, b_n, c_n, d_n) = \theta_n$ and $(a_x, b_x, c_x, d_x) = \theta_x$ depending on the state (n=normal or x=excited). The chain is set up on discrete times $0, t_1, t_2, \dots, t_{k-1}, t_k, \dots, t_{K-1}, t_K$ where $t_K = T_\alpha$ (the

expiry of the swaption). The transition matrix is:

$$\mathbf{T} = \begin{pmatrix} P_{xx} & P_{nx} \\ P_{xn} & P_{nn} \end{pmatrix} = \begin{pmatrix} P_{xx} & 1 - P_{nn} \\ 1 - P_{xx} & P_{nn} \end{pmatrix} \quad (12)$$

where P_{xn} is the probability of going from the excited to the normal state. The unconditional probabilities of being in the normal and excited state are:

$$q_n = \frac{1 - P_{xx}}{2 - P_{nn} - P_{xx}} \quad \text{and} \quad q_x = \frac{1 - P_{nn}}{2 - P_{nn} - P_{xx}} \quad (13)$$

Our intuition is that the system should spend most of its time in a normal state, and experience rare, relatively short lived transitions to the excited state. If we take the Markov step to be 1 week, then from an economic perspective we would expect P_{nn} to be close to 1.0 since the normal state is long lived, and P_{xx} to be considerable lower since the excited state is not expected to last more than a few weeks. There is one more probability to specify, which is the probability of being in the normal state at $t = 0, H_0$. Then if H_k is the probability of being in the normal state between times t_k and t_{k+1} ($\bar{H}_k = 1 - H_k$, is the probability of being in the excited state) - the state vector (\bar{H}_k, H_k) can be calculated from the transition matrix by:

$$\begin{pmatrix} \bar{H}_k \\ H_k \end{pmatrix} = \mathbf{T}^k \begin{pmatrix} \bar{H}_0 \\ H_0 \end{pmatrix}$$

Although excited states are rare, and hence have little impact on long-dated options, the probability of starting in an excited state is important because if we start in an excited state it will have a large effect on the shorter-dated options. For later convenience we define:

$$\begin{aligned} \nu_{ijk}^n &= \int_{t_k}^{t_{k+1}} \sigma_i(t; \theta_n)\sigma_j(t; \theta_n)dt \quad \text{and} \\ \nu_{ijk}^x &= \int_{t_k}^{t_{k+1}} \sigma_i(t; \theta_x)\sigma_j(t; \theta_x)dt \end{aligned} \quad (14)$$

We also define:

$$\Lambda_k^{n,\alpha\beta} = \sum_{i,j=\alpha}^{\beta-1} \zeta_i^{\alpha\beta}(0)\zeta_j^{\alpha\beta}(0)[F_i(0) + \epsilon][F_j(0) + \epsilon]\nu_{ijk}^n \quad (15)$$

and similarly for the excited state. Letting $\nu_q^{\alpha\beta} = T_\alpha(\sigma_q^{\alpha\beta})^2$ be the displaced diffusion variance for a particular Markov path, q , then one may write in a compact notation:

$$\nu_q^{\alpha\beta} = \sum_{k=0}^{K_\alpha-1} [I_k^q \Lambda_k^{n,\alpha\beta} + \bar{I}_k^q \Lambda_k^{x,\alpha\beta}] \quad (16)$$

where I_k^q is the indicator function of being in the normal state between times t_k and t_{k+1} for path q , $\bar{I}_k^q = 1 - I_k^q$ (i.e. the indicator of the excited state), and K_α is the number of Markov steps up to time T_α . So for any given Markov path the price⁵ of a swaption can easily be found.

From equation 16 it is easy to calculate expectations⁶.

$$\begin{aligned} \mathbb{E}[\nu] &= \sum_{q=1}^Q p_q \nu_q \\ &= \sum_{k=0}^{K-1} \left[\left\{ \sum_{q=1}^Q p_q I_k^q \right\} \Lambda_k^n + \left\{ \sum_{q=1}^Q p_q \bar{I}_k^q \right\} \Lambda_k^x \right] \quad (17) \\ &= \sum_{k=0}^{K-1} [H_k \Lambda_k^n + \bar{H}_k \Lambda_k^x] \end{aligned}$$

where p_q is the probability of the q^{th} Markov path, and $Q = 2^K$ which is 4.5×10^{15} for a Markov chain with a step size of 1 week and an expiry of 1 year. If $H_0 = q_n$ (i.e. we start in the unconditional state), then $H_k = q_n \forall k$, so one may write:

$$\begin{aligned} \mathbb{E}[\nu] &\equiv \hat{\nu} = q_n \sum_{k=0}^{K-1} \Lambda_k^n + q_x \sum_{k=0}^{K-1} \Lambda_k^x \\ &= T_\alpha \left\{ \begin{array}{l} q_n [\sigma_{\text{Black}}(a_n, b_n, c_n, d_n, \epsilon)]^2 + \\ q_x [\sigma_{\text{Black}}(a_x, b_x, c_x, d_x, \epsilon)]^2 \end{array} \right\} \quad (18) \end{aligned}$$

so in this case the expectation, $\mathbb{E}[\nu]$, is just the weighted sum of the Black-vol squared for the two parameter sets.

The quantity $\mathbb{E}[\nu]$ is not our finally answer, but is important for the approximation methods we adopt. We require the average of the option price over all the paths, or equivalently the implied volatility. We write the price of the swaption⁷ as $C_q = B(\nu_q)$ for a given Markov path.

⁵As is standard in the market, this article regards the Black-vol of a swaption as representing its price, because the Black[6] formula puts the two in one-to-one correspondence given that all the other inputs to the Black formula are known. Hence the volatility σ (or equivalently the variance $\nu \equiv \sigma^2 \tau$) is isomorphic with the swaption price.

⁶To prevent superscript overload we suppress α and β since they are not needed for clarity.

⁷In full, the Black price for a payers swaption, which is a call on the swap rate SR , is:

$$\begin{aligned} B(SR, X, K, T, \nu) &= A(SRN(d1) - KN(d2)), \quad (19) \\ \text{where } d1 &= \frac{\ln\left(\frac{SR}{K}\right) + \nu/2}{\sqrt{\nu}}, \end{aligned}$$

The expected price is then

$$\mathbb{E}[C] \equiv \hat{C} = \sum_{q=1}^Q p_q C_q = \sum_{q=1}^Q p_q B(\nu_q) \neq B(\hat{\nu}) \quad (20)$$

The implied volatility, $\sigma(K)$, is:

$$\sigma(K) = B^{-1}(\hat{C}) \neq \sqrt{\frac{\hat{\nu}}{T}} \quad (21)$$

Given our discrete time Markov chain, the distribution of ν is also discrete, albeit with a huge number of possible values from $\nu_{\min} \equiv T_\alpha [\sigma_{\text{Black}}(a_n, b_n, c_n, d_n, \epsilon)]^2$ to $\nu_{\max} \equiv T_\alpha [\sigma_{\text{Black}}(a_x, b_x, c_x, d_x, \epsilon)]^2$. Denoting this distribution as $f(\nu)$ we may write

$$\hat{C} = \int_{-\infty}^{\infty} B(\nu) f(\nu) d\nu = \int_{\nu_{\min}}^{\nu_{\max}} B(\nu) f(\nu) d\nu \quad (22)$$

Since we do not know directly $f(\nu)$ we must find an approximate method for calculating the integral. Given that we can easily draw random numbers from $f(\nu)$ by simulation of a Markov path, an obvious choice is Monte-Carlo. This is discussed later.

2.4 Higher Moments of ν

For a particular Markov path we have $\nu_q = \sum_{k=0}^{K_\alpha-1} [I_k^q \Lambda_k^n + \bar{I}_k^q \Lambda_k^x]$. From this we can calculate higher moments.

$$\begin{aligned} \nu_q^2 &= \left(\sum_{k=0}^{K_\alpha-1} [I_k^q \Lambda_k^n + \bar{I}_k^q \Lambda_k^x] \right) \left(\sum_{k'=0}^{K_\alpha-1} [I_{k'}^q \Lambda_{k'}^n + \bar{I}_{k'}^q \Lambda_{k'}^x] \right) \\ &= \sum_{k=0, k'=0}^{K_\alpha-1} \left(I_k^q I_{k'}^q \Lambda_k^n \Lambda_{k'}^n + \bar{I}_k^q \bar{I}_{k'}^q \Lambda_k^x \Lambda_{k'}^x + 2I_k^q \bar{I}_{k'}^q \Lambda_k^n \Lambda_{k'}^x \right) \quad (23) \end{aligned}$$

We define the following probabilities:

$$A(k - k') \equiv P(S_k = n | S_{k'} = n) = \left([\mathbf{T}^{|k-k'|}]_{11} \right) \quad (24)$$

$$B(k - k') \equiv P(S_k = x | S_{k'} = x) = \left([\mathbf{T}^{|k-k'|}]_{00} \right) \quad (25)$$

$$\bar{A}(k - k') \equiv 1 - A(k - k') \equiv P(S_k = x | S_{k'} = n) \quad (26)$$

$$\bar{B}(k - k') \equiv 1 - B(k - k') \equiv P(S_k = n | S_{k'} = x) \quad (27)$$

$d2 = d1 - \sqrt{\nu}$, and the annuity A is $A^{\alpha\beta}(t) = \sum_{i=\alpha+1}^{\beta} \tau_i P_i(t)$, where τ_i is the year fraction between t_{i-1} and t_i

where $t_k \geq t_{k'}$. The second moment is the expectation of ν^2

$$\begin{aligned} \mathbb{E}[\nu^2] &= \sum_{q=1}^Q p_q \nu_q^2 = \sum_{k=0}^{K_\alpha-1} [H_k(\Lambda_k^n)^2 + \bar{H}_k(\Lambda_k^x)^2] \\ &+ 2 \sum_{k=0, k' < k}^{K_\alpha-1} \begin{bmatrix} A(k-k')H_{k'}\Lambda_k^n\Lambda_{k'}^n \\ + B(k-k')\bar{H}_{k'}\Lambda_k^x\Lambda_{k'}^x \\ + \bar{A}(k-k')H_{k'}\Lambda_k^x\Lambda_{k'}^n \\ + \bar{B}(k-k')\bar{H}_{k'}\Lambda_k^n\Lambda_{k'}^x \end{bmatrix} \end{aligned} \quad (28)$$

For K_α Markov steps this involves $K_\alpha(K_\alpha+1)/2$ terms⁸. After some tedious manipulation the third moment can be found in the same way. The expression is:

$$\begin{aligned} \mathbb{E}[\nu^3] &= \sum_{q=1}^Q p_q \nu_q^3 = \sum_{k=0}^{K_\alpha-1} [H_k(\Lambda_k^n)^3 + \bar{H}_k(\Lambda_k^x)^3] \\ &+ 3 \sum_{\substack{k=0 \\ k' < k}}^{K_\alpha-1} \begin{bmatrix} A(k-k')H_{k'}\Lambda_k^n\Lambda_{k'}^n[\Lambda_k^n + \Lambda_{k'}^n] \\ + B(k-k')\bar{H}_{k'}\Lambda_k^x\Lambda_{k'}^x[\Lambda_k^x + \Lambda_{k'}^x] \\ + \bar{A}(k-k')H_{k'}\Lambda_k^x(\Lambda_{k'}^n)^2 + (\Lambda_k^n)^2\Lambda_{k'}^x \\ + \bar{B}(k-k')\bar{H}_{k'}[(\Lambda_k^n)^2\Lambda_k^x + \Lambda_k^n(\Lambda_{k'}^x)^2] \end{bmatrix} \\ &+ 6 \sum_{\substack{k=0 \\ k' < k \\ k'' < k'}}^{K_\alpha-1} \begin{bmatrix} A(k-k')A(k'-k'')H_{k''}\Lambda_k^n\Lambda_{k'}^n\Lambda_{k''}^n \\ + B(k-k')B(k'-k'')\bar{H}_{k''}\Lambda_k^x\Lambda_{k'}^x\Lambda_{k''}^x \\ + A(k-k')\bar{B}(k'-k'')\bar{H}_{k''}\Lambda_k^n\Lambda_{k'}^n\Lambda_{k''}^n \\ + \bar{B}(k-k')\bar{A}(k'-k'')H_{k''}\Lambda_k^x\Lambda_{k'}^x\Lambda_{k''}^x \\ + \bar{A}(k-k')A(k'-k'')H_{k''}\Lambda_k^n\Lambda_{k'}^n\Lambda_{k''}^n \\ + B(k-k')\bar{A}(k'-k'')H_{k''}\Lambda_k^x\Lambda_{k'}^x\Lambda_{k''}^n \\ + \bar{A}(k-k')\bar{B}(k'-k'')\bar{H}_{k''}\Lambda_k^n\Lambda_{k'}^n\Lambda_{k''}^x \\ + \bar{B}(k-k')B(k'-k'')\bar{H}_{k''}\Lambda_k^n\Lambda_{k'}^x\Lambda_{k''}^x \end{bmatrix} \end{aligned} \quad (29)$$

and naturally involves order K_α^3 terms.

3 Approximations to the Swap-tion Price

The expected price cannot be computed by simply adding up the contributions from the individual paths, as equation 20 suggests, since for a swaption with 10 years to expiry and a Markov period of 1 week, there are 3×10^{156} paths. In practice one does not need to consider every possible path, merely a large enough sample of *representative* paths. This is Monte Carlo pricing, and we defer the details to section 4.

⁸For a 10-year option with weekly Markov steps this is 135460 terms, which is a large computational burden.

Even the parsimonious sampling afforded by Monte Carlo can be computationally heavy however. We therefore present approximations of the swaption price that simply use the moments calculated above, and are therefore much faster than large sample Monte Carlo.

3.1 First Approximation: Taylor Series

One approximation idea is to expand the Black formula for a swaption price using a Taylor series. If we expand the swaption price about $\hat{\nu}$ rather than the usual $\hat{\sigma}$ and suppress all non-relevant indices, then:

$$\begin{aligned} B(\nu) &= B(\hat{\nu}) + (\nu - \hat{\nu}) \frac{\partial B}{\partial \nu} \Big|_{\nu=\hat{\nu}} \\ &+ \frac{(\nu - \hat{\nu})^2}{2} \frac{\partial^2 B}{\partial \nu^2} \Big|_{\nu=\hat{\nu}} + \frac{(\nu - \hat{\nu})^3}{3!} \frac{\partial^3 B}{\partial \nu^3} \Big|_{\nu=\hat{\nu}} + \dots \end{aligned} \quad (30)$$

So the integral of equation 22 can be written:

$$\begin{aligned} \hat{C} &= B(\hat{\nu}) + E[(\nu - \hat{\nu})] \frac{\partial B}{\partial \nu} \Big|_{\nu=\hat{\nu}} + \frac{E[(\nu - \hat{\nu})^2]}{2} \frac{\partial^2 B}{\partial \nu^2} \Big|_{\nu=\hat{\nu}} \\ &+ \frac{E[(\nu - \hat{\nu})^3]}{3!} \frac{\partial^3 B}{\partial \nu^3} \Big|_{\nu=\hat{\nu}} + \dots \\ &= B(\hat{\nu}) + \frac{\text{var}(\nu)}{2} \frac{\partial^2 B}{\partial \nu^2} + \frac{S(\nu)}{3!} \frac{\partial^3 B}{\partial \nu^3} + \dots \end{aligned} \quad (31)$$

Hence \hat{C} can be written as a series in the central moments with the coefficient of the n^{th} term being $\frac{1}{n!} \frac{\partial^n B}{\partial \nu^n}$.

The Taylor series expansion of the Black formula in terms of ν is however very slow to converge; $\frac{\partial^n B}{\partial \nu^n}$ has terms $\mathcal{O}(\nu^{-n+1/2})$, which means for an expansion about small ν (which is what we have for short dated options) the high order terms blow up. Cutting the series off after a small number of terms⁹ can lead to very wrong answers.

The expansion to second order works well when the variance of ν is small or $\hat{\nu}$ is large. Otherwise the first order expansion is often better in terms of how closely it agrees with a large sample Monte-Carlo. Including the third order terms can make the approximation even worse. The overall conclusion is that in order to obtain results valid under a range of conditions a better approximation technique is required.

3.2 Second Approximation: Dirac Delta

In the Taylor series approximation we may view the first order in two ways: writing $B(\nu) = B(\hat{\nu}) + (\nu - \hat{\nu}) \frac{\partial B}{\partial \nu}$;

⁹calculating moments above skew is impractical

writing $f(\nu) = \delta(\nu - \hat{\nu})$. This second case, replacing a PDF by a delta function at its mean seems crude but is equivalent to the first order Taylor expansion. We extend this and write:

$$f(\nu) \approx w\delta(\nu - \nu_1) + (1 - w)\delta(\nu - \nu_2) \quad (32)$$

i.e. two Dirac delta functions at ν_1 and ν_2 . Given the three degrees of freedom, w , ν_1 and ν_2 , we can match the mean, variance and skew of $f(\nu)$. The option price is now:

$$\hat{C} = wB(\nu_1) + (1 - w)B(\nu_2) \quad (33)$$

Perhaps surprisingly this is a very accurate approximation across all possible model parameters. To match mean, variance and skew assign the following:

$$\Delta_1 = \left(-\frac{S}{\sigma^2} + \sqrt{\frac{S^2}{\sigma^4} + 4\sigma^2}\right)/2 \quad (34)$$

$$\Delta_2 = \Delta_1 + \frac{S}{\sigma^2} \quad (35)$$

$$w = \frac{\Delta_2}{\Delta_1 + \Delta_2} \quad (36)$$

where $\Delta_1 = \hat{\nu} - \nu_1$, $\Delta_2 = \nu_2 - \hat{\nu}$ and S is the third central moment (unnormalised skew). If skew is not known, rather than set it to zero using the above equations it is better to set:

$$\Delta_1 = \frac{\hat{\nu} - \nu_{min}}{2} \quad (37)$$

$$\Delta_2 = \frac{\sigma^2}{\Delta_1} \quad (38)$$

$$w = \frac{\Delta_2}{\Delta_1 + \Delta_2} \quad (39)$$

To understand why this approximation works we Taylor expand the solution:

$$\begin{aligned} & wB(\nu_1) + (1 - w)B(\nu_2) \\ &= wB(\hat{\nu} - \Delta_1) + (1 - w)B(\hat{\nu} + \Delta_2) \\ &= B(\hat{\nu}) + (-w\Delta_1 + (1 - w)\Delta_2) \frac{\partial B}{\partial \nu} \\ &\quad + \frac{(w\Delta_1^2 + (1 - w)\Delta_2^2)}{2} \frac{\partial^2 B}{\partial \nu^2} \\ &\quad + \frac{(-w\Delta_1^3 + (1 - w)\Delta_2^3)}{3!} \frac{\partial^3 B}{\partial \nu^3} + \dots \\ &= B(\hat{\nu}) + \frac{\text{var}(\nu)}{2} \frac{\partial^2 B}{\partial \nu^2} + \frac{S(\nu)}{3!} \frac{\partial^3 B}{\partial \nu^3} + \dots \end{aligned} \quad (40)$$

where the third line follows from our choice of the values of w , Δ_1 and Δ_2 . What we have here is an implicit

Taylor series to the infinite order; the first four terms¹⁰ are exactly the same as if we had used $f(\nu)$. The higher moments are not correct but that is not so important. The Taylor approximation breaks down because we are terminating the series after only a few terms. Here the series is, in effect, infinite.

3.3 Explicit Treatment of Extreme Paths

For short dated options the probability of starting in the normal state and remaining there until expiry (or equivalently starting in the excited state) can be large. These special cases can be treated explicitly by writing:

$$f(\nu) = a\delta(\nu - \nu_{min}) + b\delta(\nu - \nu_{max}) + (1 - a - b)\tilde{f}(\nu) \quad (41)$$

where $a = H_0(P_{nn})^{K_\alpha - 1}$, $b = \bar{H}_0(P_{xx})^{K_\alpha - 1}$, and $\tilde{f}(\nu)$ is the residual probability density function. We then have:

$$\hat{C} = \int_{\nu_{min}}^{\nu_{max}} B(\nu)f(\nu)d\nu = aB(\nu_{min}) + bB(\nu_{max}) \quad (42)$$

$$+ (1 - a - b) \int_{\nu_{min}}^{\nu_{max}} B(\nu)\tilde{f}(\nu)d\nu \quad (43)$$

So we have the same type of integral to perform, albeit with a weighting factor that may be much less than one. We know the first three moments of $f(\nu)$, so the first three moments of $\tilde{f}(\nu)$ are trivially found. Now we can use the Taylor Series or Dirac Delta approximations on this residual probability density function. In the second case the call price is:

$$\begin{aligned} \hat{C} &= aB(\nu_{min}) + bB(\nu_{max}) \\ &\quad + (1 - a - b)[wB(\nu_1) + (1 - w)B(\nu_2)] \end{aligned} \quad (44)$$

where the weight w and positions ν_1 and ν_2 are chosen so the moments of the residual PDF are matched. Since equation 42 is exact this procedure will always increase the accuracy of the approximation.

3.4 Additional Approximations

If the initial probability of being in the normal state H_0 is not equal to the Unconditional probability q_n , then the probability H_k will still tend to q_n as k becomes large. Exactly how quickly this will happen, of course, depends on the transition matrix¹¹. Defining K^* as a proxy for

¹⁰including the zeroth moment.

¹¹For $H_0 = 0$ (in the excited state for sure) and $P_{nn} = 0.975$ and $P_{xx} = 0.6$, then $H_k \approx q_x = 0.0588$ after 19 steps. While for $P_{xx} = 0.95$, 140 steps are required for 4 decimal place agreement.

the step beyond which convergence is achieved, we may rewrite equation 17 as:

$$\begin{aligned} E[\nu^{\alpha\beta}] &= \frac{1}{T_\alpha} \left(\sum_{k=0}^{K^*-1} \left[H_k \Lambda_k^{n,\alpha\beta} + \bar{H}_k \Lambda_k^{x,\alpha\beta} \right] \right. \\ &\quad \left. + \sum_{k=K^*}^{K_\alpha-1} \left[q_n \Lambda_k^{n,\alpha\beta} + q_x \Lambda_k^{x,\alpha\beta} \right] \right) \\ &= \frac{1}{T_\alpha} \left(\sum_{k=0}^{K^*-1} \left[H_k \Lambda_k^{n,\alpha\beta} + \bar{H}_k \Lambda_k^{x,\alpha\beta} \right] \right) \\ &\quad + q_n \Lambda_{K^*,K_\alpha}^{n,\alpha\beta} + q_x \Lambda_{K^*,K_\alpha}^{x,\alpha\beta} \end{aligned} \quad (45)$$

where (46)

$$\Lambda_{K^*,K_\alpha}^{n,\alpha\beta} = \sum_{i,j=\alpha}^{\beta-1} \left\{ \begin{array}{l} \zeta_i^{\alpha\beta}(0) \zeta_j^{\alpha\beta}(0) [F_i(0) + \epsilon] \\ \times [F_j(0) + \epsilon] (\nu_{ij}^n(T_\alpha) - \nu_{ij}^n(T^*)) \end{array} \right\} . \quad (47)$$

If $K^* \ll K_\alpha$ this greatly reduces the number of terms that must be calculated and stored. A similar approach can be followed with the calculation of $\text{Var}(\nu^{\alpha\beta})$.

4 Comparison of the Dirac delta and Monte-Carlo Methods

As we cannot hope to enumerate every path in the Markov chain, we use large sample Monte Carlo as our benchmark against which to test other faster approximations.

Given a transition matrix \mathbf{T} and a starting probability one can simulate a set of N paths and then calculate $\nu^{\alpha\beta}$ for each path using equation 16. Our estimate of the swaption price is then

$$\widetilde{C}^{\alpha\beta} = \frac{1}{N} \sum_{i=1}^N B(SR^{\alpha\beta}, X^{\alpha\beta}, K^{\alpha\beta}, \nu_i^{\alpha\beta}) \quad (48)$$

The simulated values of ν can be adjusted so that the sample mean and variance are equal to their correct values from the moments calculated in section 2.3.

Stable Monte Carlo results can be achieved with 1000 paths, although 10,000 or more is desirable. It is not practical to calibrate the model when Monte Carlo is used for pricing. However it is sufficient to calibrate with the Dirac delta approximate and then check there is agreement with large sample Monte Carlo.

The Dirac delta method of section 3.2 was compared to results from Monte-Carlo simulation for a large sample of dates over the 8.5 year period, and in all cases good agreement was found. To illustrate the results here

we consider two dates, 3-Jan-2005 when the market was *excited*, and 3-Jan-2006 when the market was in a *normal* state. Fits were made using the Dirac delta method, then the same parameters were used to generate a sample of 100,000 paths for Monte-Carlo simulation. The forward-rate volatilities curves are shown in figure 3 and the model parameters in table 1. All the fits are performed with the constraints on the model parameters detailed in appendix B on page 15.

Figure 3: The normal and excited forward-rate displaced-volatility curves for the fits made to the USD swaption matrices on 3-Jan-2005 and 3-Jan-2006.

Figures 4 and 5 show the fit to the market using the Dirac delta method and the corresponding values using Taylor and Monte-Carlo methods with the same parameters. We are not considering the quality of the fit here, hence not showing the actual market volatilities. Our focus at this point is how well the Dirac delta and Taylor methods line up with the benchmark Monte-Carlo; in the first instances the agreement is good for both approximation methods, while in the second the shortcomings of the Taylor method are highlighted.

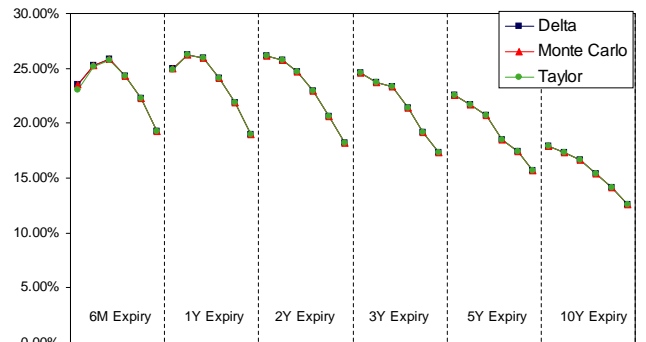


Figure 4: Fit using the Dirac delta approximation made to the 3-Jan-2005 USD swaption matrix. Here the Dirac delta lines up almost perfectly with the Monte-Carlo, and the Taylor method is slightly off on the first Swaption.

Without this accurate approximation, fitting long time series would be a massive computational burden.

	a_n	b_n	c_n	d_n	a_x	c_x	d_x	ϵ	P_{nn}	P_{xx}	\bar{H}_0
2005	-0.0012	0.0072	0.38	0.0047	0.012	0.25	0.0051	1.0	0.979	0.921	0.030
2006	-0.0044	0.0090	0.52	0.0046	0.022	3.87	0.017	1.0	0.997	0.980	0.097

Table 1: Fitted parameters for fit to USD on 3-Jan-2005 and 3-Jan-2006.

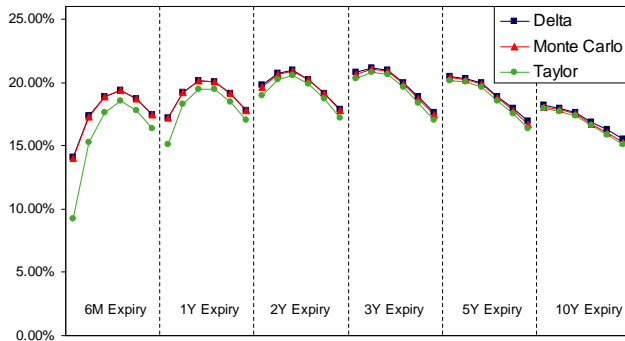


Figure 5: Fit made to the 3-Jan-2006 USD swaption matrix. Again the Dirac delta method lines up almost perfectly with the Monte-Carlo, but this time the breakdown of the Taylor approximation is clear.

5 Calibrating the model to historical data

The data sets used in this study are the Euro swaption volatilities from 16-Jan-2001 to 07-Nov-06 (1516 trading days) and the USD swaption volatilities from 5-Jan-98 to 30-Jun-2006 (2215 trading days). In both cases we consider the same 36 swaption structures which are; 6-month expiry into 1-year, 2-year, 3-year, 5-year and 10-year tenors, then 1-year, 2-year, 3-year, 5-year and 10-year expiries into the same tenors. Data for other swaption structures were available, but the points considered are the most liquid, and therefore less likely to suffer from pricing errors due to an update lag.

Figure 6 shows the evolution of volatilities for the 6-month expiry Euro swaptions. Up until late 2004 the volatilities are decreasing with swap length - this is a feature of a monotonically decreasing forward volatility curve.

Figure 7 below shows the evolution of the volatilities for the 6-month expiry and 10-year expiry USD swaptions over the 8.5 year period.

Figure 8 shows the USD swap rate over the same pe-

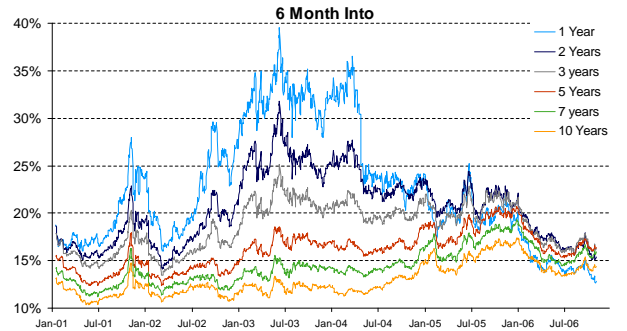


Figure 6: The evolution of the Euro swaption matrix over the period 16-Jan-01 to 07-Nov-06. Up until late 2004 the volatilities are decreasing with swap length.

Figure 7: The evolution of the USD swaption matrix over the period 5-Jan-98 to 30-Jun-2006. The very high volatility (60%) at the short end (6-month into 1-year) from mid-2002 to mid-2004 corresponds to very low short term interest rates during this period.

riod. The very high (log-normal) volatility in the swaption market corresponds to very low rates (< 2%) at the short end of the yield curve. This is the normal versus log-normal effect, because a 1% rise in interest rates when the rates are currently 2% implies a much higher log-normal volatility than the same rise when rates are at 5%.

5.1 Minimising χ^2 to fit the historical data

The entire time series was fitted using the Dirac Delta method detailed in section 3.2 by minimising the χ^2 for the difference between the market and model volatilities. χ^2 was defined by

$$\chi^2 = \sum_{i=1}^{36} \left(\frac{\sigma_i^{\text{Market}} - \sigma_i^{\text{Model}}}{\epsilon_i} \right)^2 \quad (49)$$

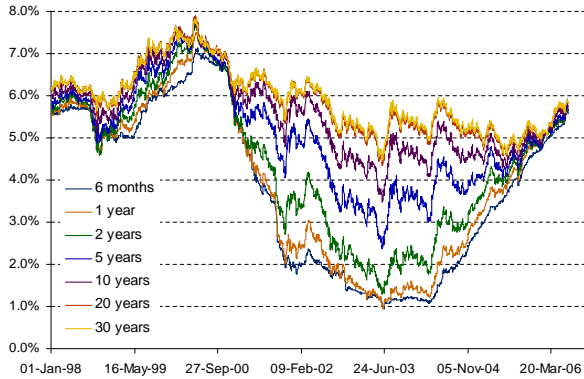


Figure 8: The evolution of the USD swap rate over the period 5-Jan-98 to 30-Jun-2006.

The error term ε_i was taken as an arbitrary value of 0.5%.

To compare volatilities between low and high interest rate environments, it is usually more convenient to work with *normal-volatility* σ_{Norm} rather than log-normal volatility σ . *Normal-volatility* is defined by $\sigma_{\text{Norm}} = SR \times \sigma$, where SR is the forward swap rate. So a swap rate of 4% with 10% log-normal volatility and a swap rate of 1% with 40% log-normal volatility both have the same *normal-volatility* of 0.40% per annum. If we define $\varepsilon_i = \text{spread}/SR_i$ for some fixed *spread*, then the chi-square above is a comparison between market and model normal-volatilities.

6 Results

6.1 Euro Market Results

The deterministic model when calibrated to the Euro data fits extremely well. All six model parameters ($a, b, c, d, \epsilon, \lambda$) have non-constant values. Figure 9 shows the normal-vol chi-square of the time series fits. We arbitrarily choose *spread* = 1bps so that the average chi-square-per-degree-of-freedom (χ^2/DoF) is close to 1. This should be compared with figure 16 which shows the results for the fit to USD market also using *spread* = 1bps.

Figures 10 and 11 show the fits with the deterministic model to the Euro market on 04-Jun-02 and 13-Jun-03 which are the best and worst fits of the series respectively. The values of the χ^2 on these dates are 25.0 and 260.0.

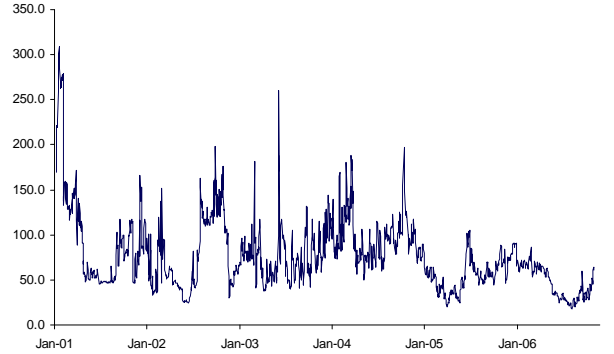


Figure 9: The normal-vol χ^2 defined in the text for the deterministic models calibrated to Euro data. The spread term is arbitrarily taken as 1bps so that average χ^2 per degree of freedom is close to 1.

Visually even the worst fit in this 6 year series is fairly good, and as we show in the next section, when compared to the same model fitted to the USD market, this simple deterministic model fits extremely well.

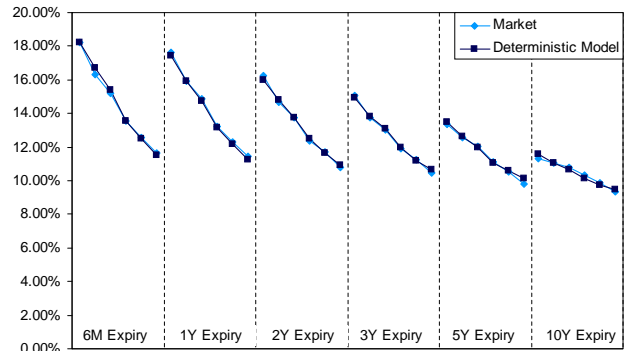


Figure 10: The Euro Swaption matrix on 04-Jun-02 and the fit with deterministic model. This is the best fit in the period.

The types of solution broadly fall into two categories depending whether the forward volatility curve is humped. If $ab > c$ then a hump exists at $\tilde{\tau} = 1/c - a/b$. Figure 12 shows the Euro market on 11-Aug-06. The fit has $\tilde{\tau} = 2.53$ years and this is reflected in the humped nature of the volatilities in common maturity buckets.

If $\tilde{\tau}$ is close to zero or no hump exists, then the volatili-

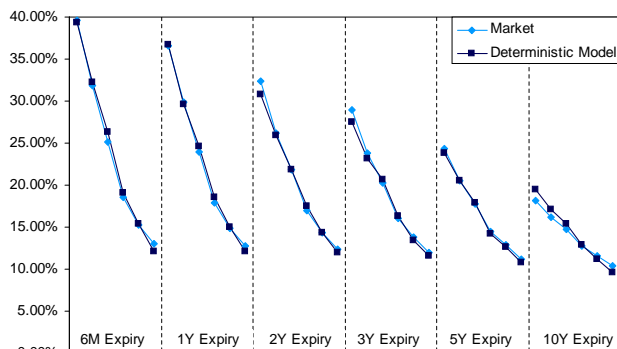


Figure 11: The Euro Swaption matrix on 16-Jun-03 and the fit with deterministic model. This is the worst fit in the period.

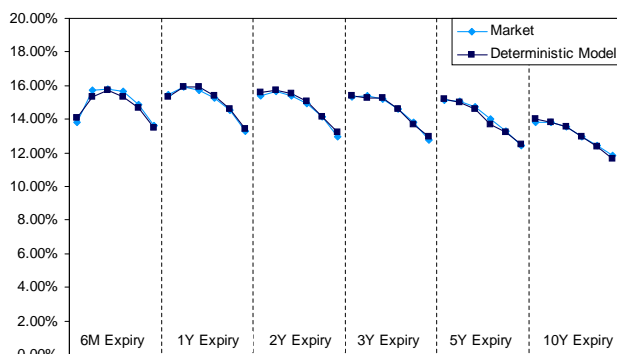


Figure 12: Euro Swaption matrix for 11-Aug-06 where the forward volatility curve has a strong humped shape. Here the hump is at $\tau = 2.53$ years.

ties in common maturity buckets are downwards slopping (in the sense of increasing swap length). An examples of this is shown in figures 10 and 11. The time series of the value of $\tilde{\tau}$, show in figure 13 is interesting - before Dec-05 the values is always less than 2 years, and monotonically decreasing volatilities are indeed observed in this period.



Figure 13: Time series of maximum of forward volatility curve, $\tilde{\tau}$, when the deterministic model is calibrated to the Euro market.

The reason the deterministic model described the Euro market so well, is that the swaption matrices seem to fall nicely into these two shapes - either the market is in a *normal* state ($\tilde{\tau} > 2$ years) and expected to stay there or the market is in an *excited* state ($\tilde{\tau} < 2$ years) and expected to stay there. The observation from the USD market, they the market can be in an *excited* state now but expected to revert to a *normal* state, is not seen in the Euro market over the period considered.

We find that the de-correlation parameter¹² λ plays an important role in the Euro market. Up until the end of Jun-01 the calibration gives $\lambda = 0$, but after this a small value is found (see figure 14). If $\lambda = 0$ is fixed throughout, the quality of the fit is significantly degraded¹³.

We mention one more parameter from the Euro market calibration - the displaced diffusion parameter ϵ . This is again interesting in that the value of ϵ suggests that the forward rates are normal until Jun-01, where they flip to being log-normal until around Sep-05, after which they start moving back towards being normal. This is show in figure 15.

¹²Recall the correlation between two forward rates i and j is given by $\rho_{ij} = e^{-\lambda|T_i - T_j|}$

¹³Although they are still good when compared to the USD fits.

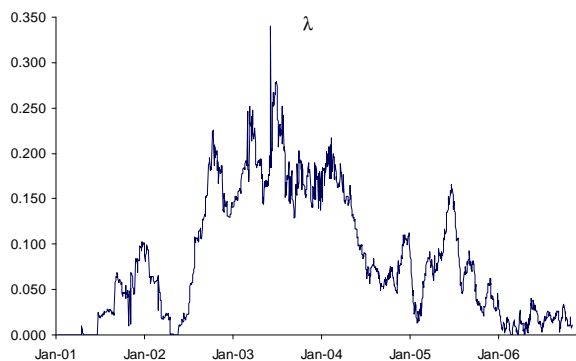


Figure 14: Time series of the de-correlation parameter λ for the calibration of the deterministic model to the Euro market

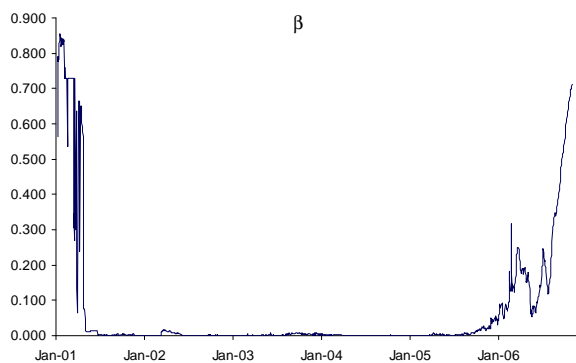


Figure 15: Time series of the displaced diffusion parameter for calibration to Euro market. Here we show $\beta = \frac{\epsilon}{1+\epsilon}$ for clarity.

The fact that such a simple model does such a good job of fitting the Euro market over a relatively long period (6 years) mean one could *parameter hedge* with the model - long/short positions such that the portfolio value is insensitive to small changes in the model parameters. Under this hedge a change in the portfolio value will only occur if a parameter moves by a large amount (i.e. second order effects are important) or the model stops being a good description of the market.

6.2 USD Results

As we alluded to in the pervious section the calibration to the USD market with the deterministic model is in general not as good, and particularly poor during periods of market excitation - the lowest χ^2 (as defined in the pervious section) is 150.9 while the largest is 11791. The corresponding values for Euro are 25.0 and 260.0, so the worst Euro fit is of comparable quality to the best USD. Two interesting observations can be made concerning the deterministic fit: the; the forward rate are all perfectly correlated. These results carry over to the Markov model, which we will consider for the rest of this section.

Figure 16 shows the χ^2 of the fit over the period for the deterministic and Markov chain models, where the error ϵ_i is defined as $\epsilon_i = \frac{1}{SR_i} \times 1bps$ so that the comparison is between the market and the model normal-volatilities. Figure 17 expands the plot for two periods of market excitation.

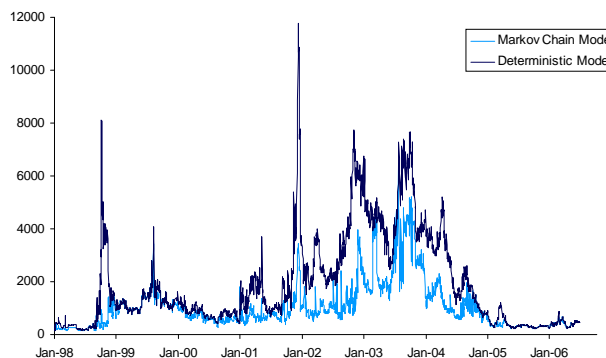


Figure 16: The χ^2 as defined in the text for the deterministic and Markov chain models calibrated to the USD market. The actual scale is somewhat arbitrary, but is the same as that used for Euro.

Figure 17: Expansion of the χ^2 around October 98 and December 2001 showing the out-performance of the Markov model.

The Markov model considerably out performs the deterministic model around October 98, December 2001 and much of the period up to September 2004. For the last 18 months both models perform equally, and this corresponds to the Markov Model starting in the normal state and remaining there.

The following two sub-sections illustrate the results obtained, to show that the results are financially plausible.

6.3 Probability of starting in the excited state

As an example of the results obtained by fitting the data by minimising χ^2 , figure 18 shows the probability of starting in the excited state at $t = 0$ (\bar{H}_0) and the unconditional probability of being in the excited state (q_x). When \bar{H}_0 is close to 1 and q_x is close to zero, the chain starts in the excited state but quickly switches to the normal state and generally stays there. The periods in October 98 and in December 2001 seen in Figure 16 when the deterministic model fails both correspond to periods when the Markov chain model starts in the excited state. Again, the period from February 2005 seems different from the previous seven years.

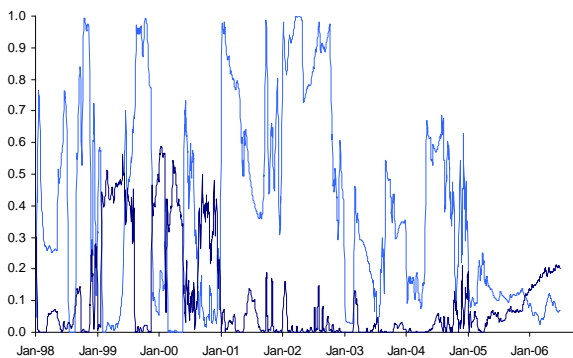


Figure 18: The probability of being in the excited state at $t = 0$ (light blue) and the unconditional probability of being in the excited state (dark blue). Shown here as a 5-day moving average.

6.4 Peak of the instantaneous volatility

Figure 19 shows the time to maturity of the peak of the instantaneous volatility in the normal state, as given by equation 3. Apart from a few dates, this is usually in the 2 ± 1 year range (to be more precise it is a mean-reverting series with a reversion level of 2.3 years). This fits in well with the qualitative justification for a humped normal volatility curve.

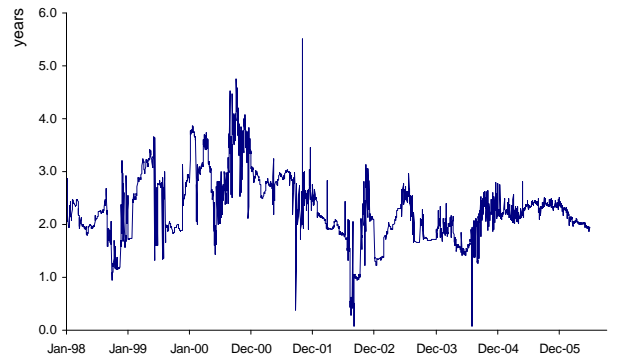


Figure 19: The time to maturity of the maximum in the normal-state volatility.

6.5 Long and short time-to-maturity instantaneous volatility

The instantaneous volatility close to expiry is given by $a + d$, while far from expiry it is given by d . Figures 20 and 21 shows the times series for $a + d$ and d for the normal volatility curve. Periods when d is zero correspond to a high probability of being in the excited state, for example the rise from zero to $\sim 0.7\%$ in mid-December 2002 corresponds to a change from the excited to the normal state. In the excited state, long dated options initially pick up their volatility from the excited curve, as they get closer to expiry the world will have reverted to a normal state, so now they ‘see’ the rising side of the normal volatility curve. The overall effect is that long dated options are fairly insensitive to whether the world is currently in an excited state.

7 Conclusion

A parsimonious model of the swaption market has been presented that consists of a hidden Markov chain with

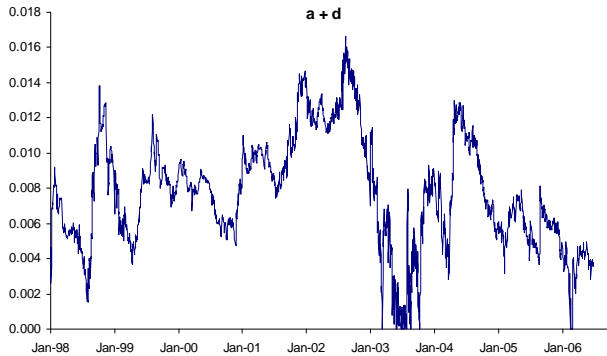


Figure 20: Time series of $a + d$ for the normal volatility curve. This is the instantaneous volatility for forward rates at maturity.

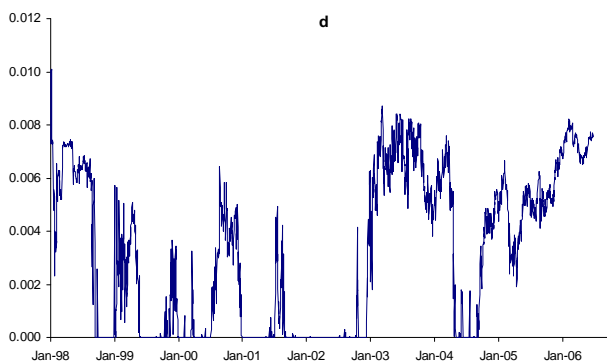


Figure 21: Time series of d for the normal volatility curve. This is the instantaneous volatility for forward rates with infinite time to maturity.

two states, normal and excited. Each state has its own curve for the instantaneous forward volatility. The volatilities in the swaption matrix result from averaging over all possible paths along the Markov chain.

A computationally fast approximation has been presented which allows the free parameters in the model to be fitted to market data. The approximation is a significant improvement on the Monte-Carlo techniques which would otherwise be necessary.

The model has been fitted successfully to the historical USD market from 1998 to mid-2006, and offers great improvement over its parent model with only one state (i.e. deterministic). We also show that the Euro market from 2001 to Nov-2006 is well described by the deterministic model - this market does not display the state changes seen in the USD market. We note, however, that the available data for EUR does not include the period of sudden market upheaval, namely Aug-Oct 1998.

A Displaced diffusion

A stochastic process $dS_t = [(1 - \beta)S_t + \beta]\sigma_t dW_t$ where $\beta \in (0, 1)$ is a martingale and log-normal for $\beta = 0$ and normal for $\beta = 1$. This can be rewritten as:

$$\frac{dS_t}{S_t + \epsilon} = \sigma_t^\epsilon dW_t \text{ where } \epsilon \equiv \frac{\beta}{1 - \beta} \text{ and } \sigma_t^\epsilon \equiv \frac{\sigma_t}{1 + \epsilon} \quad (50)$$

So $\epsilon = 0$ is log-normal and $\epsilon = \infty$ is normal; in practice the normal state is reached when $\beta \gg (1 - \beta)S_t \implies \epsilon \gg S_t$. If S_t is an interest rate ($\sim 4\%$) then the normal state is reached for any $\epsilon \gtrsim 1.0$.

The solution to equation 50 in the risk-neutral measure is:

$$S_t = (F_t + \epsilon) \exp\left(-\nu_t^\epsilon/2 + \sqrt{\nu_t^\epsilon}Z\right) - \epsilon$$

where $\nu_t^\epsilon = \int_0^t (\sigma_s^\epsilon)^2 ds$, $Z \sim N(0, 1)$ and F_t is the forward for time t . In the limit $\epsilon \rightarrow \infty$ the expression is of course:

$$S_t = F_t + \sqrt{\nu_t}Z \quad (51)$$

$$\text{where } \nu_t = (1 + \epsilon)^2 \nu_t^\epsilon = \int_0^t (\sigma_s)^2 ds \quad .$$

The price of a European call option under displaced

diffusion is

$$B(F_T, P_T, K, \nu_T^\epsilon, \epsilon) = P(0, T) \begin{bmatrix} (F + \epsilon)N(d_1) \\ -(K + \epsilon)N(d_2) \end{bmatrix}$$

$$\text{where } d_1 = \frac{\ln\left(\frac{F+\epsilon}{K+\epsilon}\right) + \nu_T^\epsilon/2}{\sqrt{\nu_T^\epsilon}}$$

$$\text{and } d_2 = d_1 - \sqrt{\nu_T^\epsilon} \quad (52)$$

and in the limit $\epsilon \rightarrow \infty$ this becomes

$$B(F_T, P_T, K, \nu_T, \epsilon) = P(0, T)\sqrt{\nu_T} \left[Z_0 N(Z_0) + \frac{\exp(-Z_0^2/2)}{\sqrt{2\pi}} \right] \quad (53)$$

$$\text{where } Z_0 = \frac{F - K}{\sqrt{\nu_T}}.$$

The *implied volatility* is found by (numerically) inverting the (non-displaced) Black formula; it is the log-normal volatility that gives the observed option price when plugged in the Black formula. Displaced diffusion always gives a volatility skew; low strike options have a higher implied log-normal volatility than high strike options. For at-the-money options, the following approximation holds

$$(\sigma_t)_{\text{Black}} \approx \frac{(F + \epsilon)\widehat{\sigma}_t^\epsilon}{F} \quad (54)$$

$$\text{where } \widehat{\sigma}_t^\epsilon = \frac{1}{t} \sqrt{\int_0^t (\sigma_s^\epsilon)^2 ds} = \sigma^\epsilon \text{ if constant}$$

This allows one to convert between displaced diffusion volatilities and log-normal volatilities which can be different orders of magnitude.

B Parameter Constraints

There are potentially 13 parameters to fit namely

- $a_n, b_n, c_n,$ and d_n describing the *normal* volatility curve,
- $a_x, b_x, c_x,$ and d_x describing the *excited* volatility curve,
- $P_{nn},$ and P_{xx} describing the Markov transition matrix,
- $\bar{H}_0,$ the probability of starting in the *excited* state,
- the displacement parameter $\epsilon,$

- the de-correlation parameter $\lambda.$

To avoid unphysical solutions, such as forward vols becoming negative or tending to infinity, and to ensure the normal and excited curves are distinct, the following constraints are imposed:

$$\begin{aligned} a_n + d_n &> 0 & b_n &> \max(0, \frac{a_n}{c_n}) \\ c_n &> 0 & d_n &> 0 \\ a_x &> 0 & b_x &= 0 \\ c_x &> 0 & d_x &> d_n + \left(\frac{b_n}{c_n} e^{-\tau c_n} - a_x e^{-\tau c_x} \right)^+ \end{aligned} \quad (55)$$

where $\tilde{\tau} = 1/c_n - a_n/b_n,$ is the time to maturity when the normal curve is at its maximum. The condition on b_n ensures that $\tilde{\tau} > 0,$ and from the fits we find that $\tilde{\tau}$ is in the range 1 to 3 years. The condition on d_x ensures that the excited curve is above the normal when the time to maturity equals 0, $\tilde{\tau}$ (the maximum of the normal curve) and infinity, although it does not ensure that the excited curve is always above the normal.

In addition to fixing $b_x = 0,$ we also fix the displacement $\epsilon = 1.0$ and the de-correlation parameter $\lambda = 0.$ So in total there are 10 free parameters to optimise.

References

- [1] Rebonato, Riccardo (2005). Forward-Rate Volatilities and the Swaption Matrix: Why Neither Time-Homogeneity Nor Time Dependence Will Do. *International Journal of Theoretical and Applied Finance*.
- [2] Peter Jaeckel & Riccardo Rebonato (2003). The Link Between Caplet and Swaption Volatilities in a BGM/Jamshidian Framework. *Journal of Computational Finance*, Vol 6, No 4, 41-59.
- [3] Damiano Brigo & Fabio Mercurio (2001). Interest Rate Models: Theory and Practice. Springer
- [4] Riccardo Rebonato (2002). Modern Pricing of Interest-Rate Derivatives: The LIBOR Market Model and Beyond. Princeton University Press.
- [5] Markowitz, Harry M. (1952). Portfolio Selection, *The Journal of Finance*, Vol .7, No. 1, pp. 77-91.
- [6] Black, Fischer (1976). The pricing of commodity contracts, *Journal of Financial Economics*, 3, 167-179.

- [7] Rubinstein, Mark (1983). Displaced Diffusion Option Pricing. *The Journal of Finance*, Vol. 38, No. 1, pp. 213-217
- [8] Rebonato R, Kainth D, (2004), A Two-Regime, Stochastic-Volatility Extension of the LIBOR Market Model. *International Journal of Theoretical and Applied Finance*, Vol 7, No 5, 555-575
- [9] Joshi M, Rebonato R, (2003), A Displaced-Diffusion Stochastic Volatility LIBOR Market Model: Motivation, Definition and Implementation, *Quantitative Finance*, Vol 3, 458-469
- [10] Marris D, (1999) 'Financial Option Pricing and Skewed Volatility', M. Phil thesis, Statistical Laboratory, University of Cambridge.
- [11] Duan J-C, Ivilina P & Ritchken P (2002). Option Pricing Under Regime Switching. *Quantitative Finance*, Vol 2, No 2, 81 - 168.
- [12] Rebonato, R. On the Simultaneous Calibration of Multifactor Lognormal Interest rate Models to Black Volatilities and to the Correlation Matrix, *Journal of Computational Finance*, 2, 4 (Summer 1999), 5-27