FAST AT-THE-MONEY CALIBRATION OF THE LIBOR MARKET MODEL THROUGH LAGRANGE MULTIPLIERS

LIXIN WU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SCIENCE AND TECHNOLOGY
HONG KONG

This version: April 3, 2002

First version: March 15, 2001

ABSTRACT. We claim to have developed the optimal methodology for non-parametric calibration of market model to the prices of at-the-money (ATM) caps/floors and swaptions, and to the historic correlations of the LIBOR rates. We take the approach of divide-and-conquer: first fit the model to historic correlations, then to the implied Black volatilities of the input options. Regularization is adopted and the calibration is cast into minimization-maximization problems by the method of Lagrange multiplier. By utilizing the quadratic functional form of both objective function and constraints, we solve the inner maximization problems with a single matrix eigenvalue decomposition, which renders the efficiency of our method. The outer minimization problems, meanwhile, are nicely subdued by gradient-based descending methods due to the convexity of the objective functions. The well-posedness of the Lagrange multiplier problems and the convergence of the descending methods are rigorously justified. Numerical results show that we have achieved very quality calibration. We have also developed a technique to calculate the hedging ratios of a derivative security with respect to the benchmark derivative instruments, using the auxiliary results of the calibration.

Key words: LIBOR market model, non-parametric calibration, constrained optimization, Lagrange multiplier method.

JEL Classification: C51, C61

Mathematics Subject Classification (2000): 60J60, 90C47

1. Introduction

The market model arose from the general framework of Heath-Jarrow-Morton (HJM, 1992) for interest-rate modeling. Rather than working on the instantaneous forward rates as in the HJM model, Brace, Gatarek and Musiela (1997), Jamshidian (1997) and Miltersen, Sandmann and Sondermann (1997) took the forward term rates as the model primitives. In their novel approach, forward term rates are assumed to follow lognormal processes under their corresponding forward measures, and, unlike the instantaneous forward rates, the log-normally diffused forward rates are non-explosive (Sandmann and Sondermann (1995)). With the log-normal forward rate processes, the use of the Black's formula (1976) by practitioners for caplets is justified. Since the new model works on the observable interest rates in the LIBOR market, it is often termed LIBOR market model.

The market model, however, does not render closed-form formula for European swaptions. Brace, Gatarek and Musiela (1997) derived an approximate formula. This formula, while being relatively complicated, was found to perform poorly whenever the reset period is a multiple of the basic period of the simple rates (Sidenius (2000)). Practitioners choice is still the Black's (1976) formula. Practically, using the Black's formula for swaptions means that the process of forward swap rates is taken to be lognormal. This is however inconsistent with the lognormality of forward rate processes for caplet pricing. Jamshidian (1997) showed that under the forward swap measure (by taking an annuity as numeraire), a forward swap rate is a martingale. Under the same measure, Andersen and Andreasen (2000) argued that the forward swap rate dynamics can be very well approximated by a lognormal process. Such approximation was supported by the empirical studies of Rebonato (2000), where he showed that the pricing errors of swaptions caused by the lognormal approximation are well within the market bid/ask spread. With these results, the simultaneous use of lognormal processes for both forward term rates and forward swap rates became well received in the market place.

Due to the multi-factor feature of market model, the valuation of exotic options with the model is typically done with Monte-Carlo simulation. Brotherton-Ratcliffe (1998) compared Euler scheme, log-Euler scheme and the 2nd order weak scheme. In terms of valuation bias, he found that log-Euler scheme is better than the Euler scheme, whilst the 2nd order scheme is superior to the log-Euler scheme. Glasserman and Zhao (1998) demonstrated that the log-Euler scheme has an inherent bias on zero-coupon bond prices, and provided alternative

simulation scheme that are free of such bias. Andersen (1999) developed the technique of early exercise boundary parameterization to valuate Bermudian options which allows exercise in a schedule of dates.

Recent research has become more focused on the correct specification of the instantaneous volatilities of the forward term rates. Non-parametric approach is preferred due to its intuitiveness and high degree of freedom. In applications, it is desirable to specify the volatilities of the forward term rates so that both the exogenously given correlation matrices and the prices of a selected set of benchmark instruments are fit. Typically, these benchmark instruments are at-the-money (ATM) caps/floors and swaptions. However, the fitting along the non-parametric approach literally brings forward a large-scale minimization problem with the number of unknowns easily reaching several hundreds. Hence it had been considered very difficult to fit both the correlations and prices for practical use. With reasonable amount of computation time Pedersen (1999) and Sidenius (2000)) have managed to achieve the calibration. To calibrate the correlations, they constructed covariance matrices and used principal component analysis to obtain the volatility components. The difficulty in such construction lies in the constraint that the ranks of the constructed covariance matrices must be less or equal to the number of factors being used. Rebonato (1999) instead constructed the low-rank approximation of the correlation matrices. To make the approximation computationally obtainable, Rebonato developed an elegant parameterization, before solving by brute force an middle scale minimization problem (with dimensions equal to the product of forward-rates number and the number of driving factors). In the literature, however, there is very little discussions on the calibration to the input swaption prices. Due to the increasing popularity of the market model in interest-rate derivative markets, the search for more efficient calibration methodology has become a very important issue.

In this paper we present extremely efficient numerical methods to accomplish the comprehensive calibration of the market models. We begin with decoupling the calibration into two sub-problems. The calibration to correlations is performed first, and the outcome is used in the calibration to the implied Black volatilities, which are equivalent to prices. The decoupled calibration problems are then formulated as constrained minimization problems and solved by the method of Lagrange multiplier. Note that by fitting to the implied Black volatilities instead of the input prices themselves, we deal with constraints as quadratic functions, as opposed to general nonlinear functions. By the Lagrange method we recast

the constrained minimization problems into minimization-maximization problems. The efficiency of our innovative approach is attributed to the facts that 1) the inner max-problem can be solved by a single matrix eigenvalue decomposition without iteration, and 2) the outer min-problem is solved conveniently by any gradient-based descending methods, due to the convexity of the objective function. As we shall see, in the environment of MATLAB, which is considered too slow for practical applications, the entire calibration for a practical problem only takes tens of seconds.

This paper is organized as follows. In §2 we introduce the background of the calibration problem and set up the corresponding mathematical formulations. In §3 we describe the methodology of our solution procedure, and offer rigorously justifications of the well-posedness of the formulations and the convergence of the numerical iteration processes. §4 contains the discussions on how to utilize the calibration output to evaluate the *deltas*, or sensitivities, of a derivative with respect the input benchmark prices. In §5 we present computational results with a practical problem. Finally in §6 we conclude.

Notation. For a square matrix A, we denote by $\mathrm{diag}(A)$ the column vector whose entries are the diagonal entries of A. Conversely, if d is a (column) vector, we define $\mathrm{diag}(d)$ the diagonal matrix whose diagonal entries are the components of d. We use $\|\cdot\|_F$ to denote the Frobenius norm for matrices and $\|\cdot\|_2$ for both the spectrum norm of a matrix and the 2-norm of a vector.

2. Problem formulation

The market model was based on the lognormal assumption of forward LIBOR rate dynamics. Let $f_j(t) = f(t; T_j, T_{j+1})$ be the arbitrage-free forward lending rate seen at time t for the period (T_j, T_{j+1}), then $f_j(t)$ is assumed to follow the lognormal process

$$df_j(t) = f_j(t)\gamma_j(t) \cdot [d\mathbf{Z}(t) - \sigma_{j+1}(t)dt],$$

where $\mathbf{Z}(t)$ is the vector of n-dimensional independent Brownian motions for some properly chosen number n, $\gamma_j(t)$ is the vector of the instantaneous volatility coefficients, and $\sigma_{j+1}(t)$ is the vector of instantaneous volatility coefficients of zero-coupon bond of maturity T_{j+1}. Consider a collection of N forward rates, $f_j, j = 1, 2, \ldots, N$. As in the Heath-Jarrow-Morton model(1992), the drifts of forward term rates in the market model are completely determined

4

by their volatilities. The no-arbitrage condition (Brace *et al.*, 1997) gives rise to their relation

$$\sigma_{j+1}(t) = -\sum_{k=1}^{j} \frac{\Delta T_k f_k(t)}{1 + \Delta T_k f_k(t)} \gamma_k(t),$$

where $\Delta T_j = T_{j+1} - T_j$ and $\gamma_j(t) = 0$ for $t \geq T_j$. As a convention we label today by $t = T_0 = 0$. The stochastic evolution of the N forward rates is fully described by the quantities of covariance defined by

$$COV_{jk}^i = \int_{T_{i-1}}^{T_i} \gamma_j(t) \cdot \gamma_k(t) dt, \quad i \leq j, k \leq N, \quad 1 \leq i \leq N.$$

Note that $COV_{jk}^i = 0$ for either $j < i$ or $k < i$ since either f_j or f_k has been reset by the time T_i. The corresponding correlations are

$$C_{jk}^i = \frac{COV_{jk}^i}{\sqrt{COV_{jj}^i} \cdot \sqrt{COV_{kk}^i}}, \quad i \leq j, k \leq N, \quad 1 \leq i \leq N.$$

For fixed i, $\{C_{jk}^i\}$ constitute an $(N - i + 1)$ by $(N - i + 1)$ non-negative symmetric matrix:

$$\mathbf{C}^i = \begin{pmatrix} C_{i,i}^i & C_{i,i+1}^i & \cdots & C_{i,N}^i \\ C_{i+1,i}^i & C_{i+1,i+1}^i & \cdots & C_{i,N}^i \\ \cdots & \cdots & \cdots & \cdots \\ C_{N,i}^i & C_{N,i+1}^i & \cdots & C_{N,N}^i \end{pmatrix}, \quad i = 1, 2, \ldots, N. \tag{1}$$

A caplet is a call option on a forward term rate. The payment of the caplet (if any) is made at the end of the forward period. Under the lognormal LIBOR rate processes, one can derive the Black's formula for, say e.g. $f_j(T_j)$, to be

$$C_j^{let}(0) = P(0, T_j)[f_j(0)N(d_1) - KN(d_2)],$$

where K is the strike rate,

$$d_1 = \frac{\ln \frac{f_j(0)}{K} + \zeta_j^2/2}{\zeta_j}, \quad d_2 = d_1 - \zeta_j,$$

and

$$\zeta_j^2 = \int_0^{T_j} \gamma_j(t) \cdot \gamma_j(t) dt = \int_0^{T_j} \|\gamma_j(t)\|^2 dt. \tag{2}$$

We call ζ_j the Black's volatility for the caplet. The Black's volatility forms a one-to-one correspondence to the caplet price. Note that in the above Black's formula we have taken the notional value to be \$1.

The lognormal forward rate process, however, does not yield closed-form solution to European swaptions. Brace *et al.* (1997) obtained an approximate formula for the swaption prices. Yet as a convention practitioners have been living with the Black's formula, partly

due to its simplicity. To justify the use of the Black's formula one must approximate a swap-rate process by a lognormal processes. Pricing consistence across the benchmarks requires the understanding of the relation between swap-rate volatilities and forward-rates volatilities. Denote

$$B^S(t) = \sum_{j=m}^{n-1} \Delta T_j P(t, T_{j+1}),$$

where $P(t, T_{j+1})$ is the time-t price of the zero-coupon bond paying \$1 at maturity T_{j+1}. Then the fair swap rate for the period (T_m, T_n) seen at time t is defined by

$$R_{m,n}(t) = \frac{P(t, T_m) - P(t, T_n)}{\sum_{j=m}^{n-1} \Delta T_j P(t, T_{j+1})}.$$

Jamshidian (1997) showed that, by Ito's lemma, the dynamics of $R_{m,n}(t)$ is

$$dR_{m,n}(t) = \sum_{j=m}^{n-1} \frac{\partial R_{m,n}(t)}{\partial f_j(t)} f_j(t) \gamma_j(t) \cdot dW^S(t), \tag{3}$$

where $W^S(t)$ is the n-dimensional independent Brownian motions under the *forward swap measure* induced by using $B^S(t)$ as the numeraire, while

$$\frac{\partial R_{m,n}(t)}{\partial f_j(t)} = \frac{\Delta T_j R_{m,n}(t)}{1 + \Delta T_j f_j(t)} \left[\frac{P(t, T_n)}{P(t, T_m) - P(t, T_n)} + \frac{\sum_{k=j}^{n-1} \Delta T_k P(t, T_{k+1})}{B^S(t)} \right].$$

The swap rate process (3) is apparently not lognormal. Yet based on (3) Anderson and Andreasen (2000) proposed the following "frozen coefficient" approximation:

$$dR_{m,n}(u) = R_{m,n}(u) \sum_{j=m}^{n-1} w_j(t) \gamma_j(u) \cdot dW^S(u), \quad t \le u < T_m, \tag{4}$$

where

$$w_j(t) = \frac{\partial R_{m,n}(t)}{\partial f_j(t)} \frac{f_j(t)}{R_{m,n}(t)}. \tag{5}$$

The above approximation was based on the observation that $w_j(u), m \le j \le n-1$ usually demonstrate negligible variation. There are other approximations, yet (4,5) appear more appealing because the coefficients are frozen after the application of the Ito's lemma. The lognormal dynamics (4,5) gives rise again to the Black's formula for European swaptions

$$PS(t, T_m, T_n) = B^S(t)[R_{m,n}(t)N(g_+) - KN(g_-)], \tag{6}$$

where K is the strike of a swaption, and

$$g_+ = \frac{\ln \frac{R_{m,n}(t)}{K} + \frac{1}{2}\zeta_{m,n}^2(t)}{\zeta_{m,n}(t)}, \tag{7}$$

$$g_- = g_+ - \zeta_{m,n}(t),$$

6

and

$$\zeta_{m,n}^2(t) = \int_t^{T_m} \left\| \sum_{j=m}^{n-1} w_j(t)\gamma_j(s) \right\|^2 ds$$

$$= \sum_{j=m}^{n-1} \sum_{k=m}^{n-1} w_j(t)w_k(t) \int_t^{T_m} \gamma_j(s) \cdot \gamma_k(s)ds. \tag{8}$$

Note that a caplet is a special case of swaption, corresponding to $n = m+1$. Hence formulae (6-8) apply to both caplets and swaptions.

In this paper, we consider the following comprehensive calibration problem that, given the implied Black's volatilities, $\{\zeta_j\}$ and $\{\zeta_{m,n}\}$, of the ATM caplets and swaptions and correlation matrices $\{\mathbf{C}^i\}$ for the rates, determine from equations (1), (2) and (8) the implied volatility functions $\gamma_j(t), j = 1, \dots, N$. We will take the non-parametric approach and look for the volatilities in the form of piece-wise constant function in t:

$$\gamma_j(t) = \gamma_j^i = s_j^i(a_{j,1}^i, a_{j,2}^i, \dots, a_{j,n}^i) \equiv s_j^i \mathbf{a}_j^i, \qquad \text{for} \quad T_{i-1} \leq t \leq T_i, \quad i \leq j,$$

with

$$s_j^i = \|\gamma_j^i\|_2, \quad \text{and} \quad \|\mathbf{a}_j^i\|_2 = 1. \tag{9}$$

It will become clear shortly why we single out s_i^j, the norm of γ_j^i, as another variable. The total number of unknowns is proportional to $n \times N^2$, which in practice can be in the magnitude of hundreds and far bigger than the number of input prices and elements of correlation matrices. Hence, we are facing a middle- to large-scale under-determined problems. Fortunately, the determination of $\{\mathbf{a}_j^i\}$ depends only on the input correlations $\{\mathbf{C}^i\}$. To see this, suppose the rank of $\mathbf{C}^i$ is less than or equal to n. Perform eigenvalue decomposition on $\mathbf{C}^i$:

$$\mathbf{C}^i = U\Lambda U^T, \tag{10}$$

where Λ is an n by n diagonal matrix with non-negative diagonal elements, and define $\mathbf{a}_j^i$ as the jth row of $U\Lambda^{1/2}$, i.e.,

$$\mathbf{a}_j^i = e_j^T U\Lambda^{1/2}, \tag{11}$$

then we have

$$\mathbf{C}^i = \begin{pmatrix} \mathbf{a}_i^i, \\ \vdots \\ \mathbf{a}_N^i \end{pmatrix} \left((\mathbf{a}_i^i)^T, \dots, (\mathbf{a}_N^i)^T \right). \tag{12}$$

7

By (9) and (12), the forward-rate dynamics in discrete form reads

$$\Delta f_j(t_i) = f_j(t_i)s_j^i \mathbf{a}_j^i \cdot [\Delta Z(t_i) - \sigma_{j+1}\Delta t_i]$$

for some small Δt_i, and the model correlation is given by

$$Corr(\Delta f_j(t_i), \Delta f_k(t_i)) = \frac{\Delta t_i \mathbf{a}_j^i \cdot \mathbf{a}_k^i}{\sqrt{\Delta t_i}\|\mathbf{a}_j^i\|_2 \cdot \sqrt{\Delta t_i}\|\mathbf{a}_k^i\|_2} = C_{jk}^i$$

Note that the columns of matrix $U\Lambda^{1/2}$ are called *principal components* of the matrix $\mathbf{C}^i$.

The complication in the determination of $\mathbf{a}_j^i, j = 1, \dots, n$ is that matrix $\mathbf{C}^i$ in general has full rank, which is equal to $N - i + 1$ and can be much bigger than n, the number of forward rate "alive". In such case the above procedure for calculating $\mathbf{a}_j^i, j = 1, \dots, n$ breaks down. Therefore, a preprocessing is in general needed to reduce the ranks of the given correlation matrices. For a given correlation matrix $\mathbf{C}^i$, we formulate the preprocessing as the following minimization problem with constraints:

$$\min_{\hat{\mathbf{C}}^i} \|\mathbf{C}^i - \hat{\mathbf{C}}^i\|_F,$$
$$\text{s.t.} \quad \hat{\mathbf{C}}^i \geq 0, \quad \text{rank}(\hat{\mathbf{C}}^i) \leq n, \quad \hat{C}_{kk}^i = 1, \quad k = i, \dots, N. \tag{13}$$

where $\hat{C} \geq 0$ means that $\hat{C}$ is a non-negative matrix. Once we have obtained $\hat{C}^i$, $\mathbf{a}_j^i$ follows from (10,11). Presumably we need to solve (13) for $i = 1, 2, \dots, N$.

Having obtained the low-rank approximation of the correlation matrices, we can proceed to the determination of the forward rate volatilities s_j^i, subject to the input prices. The number of $\{s_j^i\}$ to be determined is $N(N+1)/2$. This is typically much higher than the number of the input prices. Hence, this is an under determined problem and regularization must be adopted to make the problem well-posed. For uniqueness and smoothness of the volatility surface, we consider the following objective function

$$\|\nabla \mathbf{s}\|^2 + \epsilon\|\mathbf{s} - \mathbf{s}_0\|^2 \equiv -(\mathbf{s}, (\nabla \cdot \nabla)\mathbf{s}) + \epsilon\|\mathbf{s} - \mathbf{s}_0\|^2 \quad \text{for some } \epsilon > 0, \tag{14}$$

where $(\nabla \cdot \nabla)$ stands for the discrete Laplacian:

$$(\nabla \cdot \nabla)s_j^i = s_{j-1}^i + s_{j+1}^i + s_j^{i-1} + s_j^{i+1} - 4s_j^i. \tag{15}$$

In details (14) reads

$$\sum_{i=1}^N \sum_{j=i}^N s_j^i \left(-s_{j-1}^i - s_{j+1}^i - s_j^{i-1} - s_j^{i+1} + 4s_j^i\right) + \epsilon \sum_{i=1}^N \sum_{j=i}^N (s_j^i - s_{j,0}^i)^2. \tag{16}$$

Note that $\mathbf{s}_0$ is a *priori* volatility surface prescribed as, say for example, the volatility surface of the previous day. Hence by minimizing (14), we achieve a balance between the smoothness and the stability of the volatility surface. In (16), there are some "ghost" variables whose

sub- or sup-indeces are out of the designated range. These "ghost" variables are eliminated by using "Neumann boundary condition"[1]:

$$s_j^0 = s_j^1, \qquad j = 1, 2, \dots, N,$$

$$s_{N+1}^i = s_N^i,$$

$$s_i^{i+1} = s_i^i, \qquad i = 1, 2, \dots, N,$$

$$s_{i-1}^i = s_i^i.$$

Augmented to (14) are constraints in terms of the implied Black's volatilities:

$$\zeta_j^2 = \sum_{i=1}^{j} \Delta T_{i-1}(s_j^i)^2, \quad \text{for some } j, \tag{17}$$

for caplets and

$$\zeta_{m,n}^2 = \sum_{j,k=m,n-1} w_j w_k \sum_{i=1}^{m} s_j^i s_k^i \hat{C}_{j,k}^i \Delta T_{i-1}$$

$$= \sum_{i=1}^{m} \Delta T_{i-1} \sum_{j,k=m,n-1} s_j^i s_k^i (w_j w_k \hat{C}_{j,k}^i), \quad \text{for some } m \text{ and } n, \tag{18}$$

for swaptions. Here in (18) $\hat{C}_{jk}^i$ is an element of $\hat{\mathbf{C}}^i$, the low-rank approximation to correlation matrix $\mathbf{C}^i$. When $w_j = 1, w_k = 0, k \neq j$, condition (18) reduces to (17). Note that all functions in (16)-(18) are quadratic functions in $\{s_j^i\}$. Such feature, as we shall see later, gives rise to a powerful numerical method.

For efficiency, we will take advantage of matrix operations. For this purpose we rewrite the problem with matrix notations. First we line up the volatilities in a one-dimensional array

$$\mathbf{X} = \begin{pmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \vdots \\ \mathbf{s}^N \end{pmatrix}$$

with

$$\mathbf{s}^i = \begin{pmatrix} s_i^i \\ s_{i+1}^i \\ \vdots \\ s_N^i \end{pmatrix}.$$

We then define the matrix corresponding to the discrete Laplacian in (15) as

$$B = diag(-1, -1, 4, -1, -1).$$

---

[1]Corresponding to the boundary condition of zero normal derivative for partial differential equations. The application here may result in multiple definitions of "ghost" values beyond the range of the indeces, which is however, harmless.

Finally we associate each instrument with the following "weight" matrix

$$W_{m,n} = diag(0, \ldots , 0, w_m, \ldots , w_{n-1}, 0, \ldots , 0),$$

and "correlation matrix"

$$G_{m,n} = diag(\Delta T_0 W_{m,n} \hat{C}^1 W_{m,n}, \Delta T_1 W_{m,n} \hat{C}^2 W_{m,n}, \ldots , \Delta T_{m-1} W_{m,n} \hat{C}^m W_{m,n}, 0, \ldots , 0).$$

With the above matrices, the calibration to prices under the objective function (14) can be cast into a "non-standard" problem of quadratic programming:

$$
\begin{aligned}
&\min_X X^T B X + \epsilon(X - X_0)^T(X - X_0), \\
&\text{s.t.} \quad X^T G_{m,n} X = \zeta_{m,n}^2 \quad \text{for some } m \text{ and } n.
\end{aligned}
\tag{19}
$$

The objective function in (19) can be simplified further. Expanding the function we have

$$
\begin{aligned}
&X^T B X + \epsilon(X - X_0)^T(X - X_0) \\
={}&(X + \epsilon(B + \epsilon I)^{-1} X_0)^T(B + \epsilon I)(X + \epsilon(B + \epsilon I)^{-1} X_0) \\
&+ \epsilon(X_0)^T(I + (B + \epsilon I)^{-1})X_0,
\end{aligned}
$$

where the last term is a constant and thus can be ignored for optimization purpose. Denote

$$A = B + \epsilon I,$$

which is a positive-definite matrix, and

$$\tilde{X}_0 = -\epsilon(B + \epsilon I)^{-1} X_0,$$

we end up with the following formulation for price calibration

$$
\begin{aligned}
&\min_X (X - \tilde{X}_0)^T A (X - \tilde{X}_0), \\
&\text{s.t.} \quad X^T G_{m,n} X = \zeta_{m,n}^2 \quad \text{for some } m \text{ and } n.
\end{aligned}
\tag{20}
$$

The recast form (20) highlights that the objection function is in quadratic form with a positive definite matrix. Nevertheless, (20) can not be solved literally by Lagrange method because for an arbitrary set of finite multipliers, the Lagrange function (to be defined shortly) may not have finite minimum or maximum. For this reason, we superimpose a convex function to the objective function:

$$
\begin{aligned}
&\min_X U\left((X - \tilde{X}_0)^T A (X - \tilde{X}_0)\right), \\
&\text{s.t.} \quad X^T G_{m,n} X = \zeta_{m,n}^2 \quad \text{for some } m \text{ and } n,
\end{aligned}
\tag{21}
$$

where $U(y)$ is superlinear and monotonically increasing function for $y \geq 0$, for examples, $U(y) = y^2, y \ln y$ or e^y. Problem (21) then shares the same constrained minimum(s) with (20). We will proceed next to solve the problem with the Lagrange method.

3. Solution Methodology

In this section, we will develop numerical methods to solve the constrained minimization problems (13) and (21). The methodology to be taken is the combinations of the methods of Lagrange multiplier and steepest descend. In developing the numerical methods, we have taken full advantages of the special structure of the objective functions and constraints. Moreover, we have justified rigorously the well-posedness of the Lagrange multiplier problems and obtained the convergence of the numerical methods.

3.1. **Eigen-decomposition-based rank reduction algorithm.**

For a given non-negative symmetric N by N matrix C, we define a low-rank approximation as the solution to the following problem

$$
\min_X \|C - X\|_F,
$$
$$
\text{s.t.} \quad \text{rank}(X) \le n < N, \quad \text{diag}(X) = \text{diag}(C). \tag{22}
$$

We denote the feasible set of solutions by

$$
\mathcal{F} = \{\, X \in \mathcal{R}^{N \times N} \;\big|\; \text{rank}(X) \le n, \ \text{diag}(X) = \text{diag}(C) \,\},
$$

and any solution to problem (22) by C^*. For applications in the market model, C^* will serve subsequently as a correlation matrix and thus is expected to be a non-negative definite matrix. It may seem that the feasible set of the optimal problem should be $\mathcal{F}^+$, the subset of $\mathcal{F}$ that consists of only positive semi-definite matrices. Yet it proved recently by Zhang and Wu (2001) that the solutions to (22) are automatically positive semi-definite. Hence the explicit imposition of the extra constraint becomes unnecessary.

Following the general approach of Lagrange method, we transform the above constrained minimization problem into an equivalently min-max problem. Let $\mathcal{R}_n$ be the set of $N \times N$ matrices with rank less or equal to n. The Lagrange multiplier problem corresponding to (22) is defined as:

$$
\min_d \max_{X \in \mathcal{R}_n} L(X, d), \tag{23}
$$

with the Lagrange function

$$
L(X, d) = - \|C - X\|_F^2 - 2d^T \text{diag}(C - X), \tag{24}
$$

where d is the vector of the multipliers. Note that $L(X, d)$ is linear in d in the following sense:

$$
L(X, td + (1-t)\hat{d}) = tL(X, d) + (1-t)L(X, \hat{d}). \tag{25}
$$

11

We will rigorously justify later that the min-max problem (23) is equivalent to the original problem (22).

In numerical implementation the min-max problem (23) is solved as an minimization problem of the form

$$\min_{d} V(d), \tag{26}$$

with the objective function defined by

$$V(d) = \max_{X \in \mathcal{R}_n} L(X, d). \tag{27}$$

Hence it is a matter to find efficient methods separately for the maximization problem (27) and minimization problem (26).

For the maximization problem (27), it is crucial to observe that the Lagrange function can be written into

$$L(X, d) = -\|C + D - X\|_F^2 + \|d\|_2^2, \tag{28}$$

where D is the diagonalized matrix of d: $D = \operatorname{diag}(d)$. For given d, the maximizer to (26) can be obtained by the eigenvalue decomposition of matrix $C + D$ (which is symmetric). Let

$$C + D = U \Lambda U^T \tag{29}$$

be the eigenvalue decomposition with orthogonal matrix U and eigenvalue matrix

$$\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N),$$

where the diagonal elements are put in the decreasing order in magnitude:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|.$$

Note that both U and Λ depend on the multiplier vector d, hence they will also be denoted by $U(d)$ and $\Lambda(d)$ when highlighting the dependence is necessary. The solutions to the problem (27), the best rank-n approximations of $C + D$, are obviously given by

$$C(d) \equiv C_n(d) = U_n \Lambda_n U_n^T, \tag{30}$$

where U_n is the matrix consisting of the first n columns of U, and $\Lambda_n = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the principal submatrix of Λ of degree n. Consequently we have

$$V(d) = -\sum_{j=n+1}^{N} \lambda_j^2 + \|d\|_2^2. \tag{31}$$

Clearly, when $|\lambda_n| > |\lambda_{n+1}|$, the solution to (27) is unique. In the case $|\lambda_n| = |\lambda_{n+1}|$, the solutions become nonunique. The complication for the case of $|\lambda_n| = |\lambda_{n+1}|$ has been

12

studied by Zhang and Wu (2001). Throughout this paper we limit ourselves to the case of $|\lambda_n| > |\lambda_{n+1}|$.

Regarding the existence of solution(s) of the min-max problem (26,27), we have

Theorem 3.1. *There exists at least one solution to (26,27), and any local minimum to (26,27) is also a global minimum.*

PROOF: To prove existence we use the method of contradiction. Suppose there is no solution to (26,27), then there exists a sequence of $d^{(j)} = \{d_i^{(j)}\} \rightarrow \infty$ while $V(d^{(j)})$ decreases. Write $D^{(j)} = \text{diag}(d^{(j)}) = D_1^{(j)} + D_2^{(j)}$ as a direct sum of two diagonal matrices with $\text{rank}(D_1^{(j)}) \leq n$ and $\|D_1^{(j)}\|_\infty = \|D^{(j)}\|_\infty \longrightarrow +\infty$. Since

$$V(d^{(j)}) = \max_{X \in \mathcal{R}_n} L(X, d^{(j)}) \geq L(D_1^{(j)}, d^{(j)}),$$

we have, for $L(X, d)$ in the form of (28),

$$
\begin{aligned}
V(d^{(j)}) &\geq -\|C + D_2^{(j)}\|_F^2 + \|D^{(j)}\|_F^2 \\
&= -\|C\|_F^2 - 2tr(CD_2^{(j)}) + \|D_1^{(j)}\|_F^2 \\
&\geq -\|C\|_F^2 + \left(\|D^{(j)}\|_\infty - 2tr(|C|)\right)\|D^{(j)}\|_\infty \longrightarrow +\infty,
\end{aligned}
$$

which contradicts to the assumption that $V(d^{(j)})$ decreases. Here $|C|$ denotes the matrix with entries of C in absolute values. The existence of solution(s) follows.

The property that any local minimum must be at the same time a global minimum is due to the convexity of $V(d)$. To see this, we consider any two points $d^{(1)}$ and $d^{(2)}$. For any $t \in (0, 1)$ we have

$$
\begin{aligned}
V(td^{(1)} + (1-t)d^{(2)}) &= \max_X -\|C - X\|_F^2 - 2(td^{(1)} + (1-t)d^{(2)})^T\text{diag}(C - X) \\
&= \max_X t\left(-\|C - X\|_F^2 - 2(d^{(1)})^T\text{diag}(C - X)\right) \\
&\quad + (1-t)\left(-\|C - X\|_F^2 - 2(d^{(2)})^T\text{diag}(C - X)\right) \\
&\leq t\max_X -\|C - X\|_F^2 - 2(d^{(1)})^T\text{diag}(C - X) \\
&\quad + (1-t)\max_X -\|C - X\|_F^2 - 2(d^{(2)})^T\text{diag}(C - X) \\
&= tV(d^{(1)}) + (1-t)V(d^{(2)}).
\end{aligned}
$$

Suppose there are two local minimums d^* and d^{**} such that $V(d^*) > V(d^{**})$, we consider $d(t) = td^* + (1-t)d^{**}$ for $t \in (0,1)$. By the convexity we have

$$
\begin{aligned}
V(d(t)) &= V(td^* + (1-t)d^{**}) \\
&\leq tV(d^*) + (1-t)V(d^{**}) \\
&< tV(d^*) + (1-t)V(d^{**}) = V(d^*).
\end{aligned}
\tag{32}
$$

Let t approach 1, then $d(t)$ approaches d^*, and the inequality in (32) is in contradiction to the assumption that d^* is a local minimum. Hence we arrive at the second conclusion $\square$.

For the analytical properties of the functions involved in the solution of (27) we have

Theorem 3.2. *When $|\lambda_n(C+D)| > |\lambda_{n+1}(C+D)|$,*

(1) *the optimal solution $C(d)$ to (27) is unique and differentiable in d;*

(2) *$V(d)$ is second-order continuously differentiable in a neighborhood of d;*

(3) *if $\nabla V(d) = 0$, then d must be a global minimizer of $V(d)$.*

PROOF: The uniqueness is obtained by construction, as all solutions must be in the form (30), where Λ_n is unique when $|\lambda_n(C+D)| > |\lambda_{n+1}(C+D)|$, so is $C(d)$. Since C is symmetric and D is only a diagonal matrix, Gerschgorin Theorem in linear algebra (see for instance, Steward and Sun (1990)) implies that all eigenvalues (not necessarily distinct from each other) and eigenvectors are differentiable in a neighborhood of d, so are $V(d)$ and $C(d)$. The partial derivatives of $V(d)$ and $C(d)$ are related by the chain rule:

$$
\frac{\partial V(d)}{\partial d_k} = \sum_{ij} \frac{\partial L(C(d), d)}{\partial X_{ij}} \frac{\partial X_{ij}}{\partial d_k} + \frac{\partial L(C(d), d)}{\partial d_k}, \quad 1 \leq k \leq N.
$$

For fixed d, the optimality of $C(d)$ implies that

$$
\frac{\partial L(C(d), d)}{\partial X_{ij}} = 0, \qquad \text{for all } i \text{ and } j.
\tag{33}
$$

Consequently we have

$$
\frac{\partial V(d)}{\partial d_k} = -2(C_{kk} - C_{kk}(d)).
\tag{34}
$$

Differentiating (34) with respect to d_l yields

$$
\frac{\partial^2 V(d)}{\partial d_k \partial d_l} = 2 \frac{\partial C_{kk}(d)}{\partial d_l}, \quad \forall k \text{ and } l,
$$

whose continuity follows from that of $\frac{\partial C_{kk}(d)}{\partial d_l}$.

Next we will show that any critical points must be the global minimum. In fact if $\nabla V(d) = 0$ but d is not a global minimum, we must have another point $\hat{d}$ such that

14

$V(\hat{d}) < V(d)$. From convexity property we have

$$V(t\hat{d} + (1 - t)d) \leq tV(\hat{d}) + (1 - t)V(d), \qquad \text{for any } t \in (0, 1).$$

The above equation can be rewritten into

$$\frac{V(t\hat{d} + (1 - t)d) - V(d)}{t} \leq V(\hat{d}) - V(d) < 0.$$

Let $t \to 0$ we than have

$$\frac{(\hat{d} - d)}{\|\hat{d} - d\|_2} \cdot \nabla_d V(d) \leq V(\hat{d}) - V(d) < 0, \tag{35}$$

which is contradict to the zero-gradient condition $\nabla_d V(d) = 0$ for critical point d. The lemma is thus proved $\square$

Based on the existence of minimizer and differentiability of the value function, we can establish the equivalence between the constrained minimization problem (22) and the Lagrange multiplier problem (23).

Theorem 3.3. *Let d^* be any minimizer of the Lagrange multiplier problem (23). If $|\lambda_n(C + D^*)| > |\lambda_{n+1}(C + D^*)|$ and $\text{diag}(C) > 0$, then d^* is the unique minimizer and $C(d^*)$ solves the constrained minimization problem (22).*

PROOF: Assume to the contrary that there exists $d^{**} \neq d^*$ such that $V(d^{**}) = V(d^*)$. Denote $d(t) = d^* + t(d^{**} - d^*)$. The convexity property of $V(d)$ yields $V(d(t)) = V(d^*)$. Denote $C(t)$ as $C(d(t))$. Due to the linearity of $L(X, d)$ in d, we have that for $t \in [0, 1]$

$$V(d^*) = V(d(t)) = L(C(t), d(t))$$

$$= (1 - t)L(C(t), d^*) + tL(C(t), d^{**})$$

$$\leq (1 - t)\max_{X \in \mathcal{F}} L(X, d^*) + t\max_{X \in \mathcal{F}} L(X, d^{**})$$

$$= (1 - t)V(d^*) + tV(d^{**}) = V(d^*).$$

From the above equalities we get $L(C(t), d^*) = V(d^*)$, or

$$\|C + D^* - C(t)\|_F = \min_{X \in \mathcal{R}_n} \|C + D^* - X\|_F.$$

It follows that, by the uniqueness of $C(d^*)$,

$$C(t) = C(0), \quad t \in [0, 1]$$

is the optimal solution corresponding to $d(t)$. The eigen-decomposition of matrix $C + D(t)$ is then

$$C + D(t) = U(t)\Lambda(t)U(t)^T = C(0) + E(t),$$

where
$$C(0) = U_n(0)\Lambda_n(0)U_n^T(0),$$
$$E(t) = G_n(t)\Theta_n(t)G_n^T(t),$$

and $G_n(t)$ consists of the last $N - n$ columns of $U(t)$. Since the columns of $G_n(0)$ and $G_n(t)$ form two orthogonal bases of the null space of $C(0)$, there exists an orthogonal matrix $W(t)$ such that $G_n(t) = G_n(0)W(t)$, and thus
$$E(t) = G_n(0) * (W(t)\Theta_n(t)W^T(t)G_n(0).$$

Substituting the above expression into the equality
$$E(t) - E(0) = t(D - D^*)$$

we obtain
$$t(D - D^*) = G_n(0) * (W(t)\Theta_n(t)W^T(t) - \Theta_n) * G_n^T(0),$$

or
$$(D - D^*) = G_n(0) * H(t) * G_n^T(0) \quad \text{for some } H(t).$$

Post-multiplying $U_n(0)$ to the above equation and recalling the orthogonality between $G_n(0)$ and $U_n(0)$, we end up with
$$(D - D^*)U_n(0) = 0.$$

Clearly, if the ith diagonal of $D - D^*$ is not zero, then the ith row of $U_n(0)$ must be zero, which in turn implies the ith row of $C(0)$, including the diagonal $C_{ii}(0)$, vanishes. This is however contradict to the assumption $diag(C(d)) = diag(C) > 0$ and hence there can not be more than one minimizer.

If d^* solve the min-max problem (23), then, due to the differentiability of $V(d)$, d^* must be a critical point of $V(d)$ and its gradient vanishes, that is,
$$0 = \frac{\partial V(d)}{\partial d_k} = -2(C_{kk} - C_{kk}(d)), \quad 1 \leq k \leq N$$

from Theorem 3.2. Hence for any other matrix $\tilde{C} \in \mathcal{F}$ we have
$$V(d^*) = -\|C - C(d^*)\|_F^2$$
$$= \max_{X \in \mathcal{F}} -\|C - X\|_F^2$$
$$\geq -\|C - \tilde{C}\|_F^2,$$

meaning that $C(d^*)$ is the only solution to the constrained minimization problem (23) $\quad \square$

We have shown that the inner maximization problem (27) can be solved nicely by a single eigenvalue decomposition. The outer minimization (26) then is dealt with the method

16

of steepest descend. The convexity of $V(d)$ renders efficiency of the descending method. The algorithm is described below.

Algorithm: Take $D^{(0)}$ to be a null matrix, and repeat the following steps:

(1) Compute the eigen-decomposition of $C + D^{(k)}$: $C + D^{(k)} = U^{(k)}\Lambda^{(k)}(U^{(k)})^T$; set $\alpha^{(k)} = 1$ and $\nabla V(d^{(k)}) = -2\mathrm{diag}(C - U_n^{(k)}\Lambda_n^{(k)}(U_n^{(k)})^T)$;

(2) Define $d^{(k+1)} = d^{(k)} - \alpha^{(k)}\nabla V(d^{(k)})$;

(3) If $V(d^{(k+1)}) > V(d^{(k)}) - \frac{\alpha^{(k)}}{2}\|\nabla V(d^{(k)})\|^2$, take $\alpha^{(k)} := \alpha^{(k)}/2$, go back to step 2;

(4) if $\|d^{(k+1)} - d^{(k)}\|_2 > tol$, go to step 1;

(5) Take $d^* = d^{(k+1)}$ and $C^* = U_n^{(k)}\Lambda_n^{(k)}(U_n^{(k)})^T$.

For the convergence of the descending method we have

Theorem 3.4. *The sequence $\{d^{(k)}\}$ is bounded and hence accumulation points of $\{d^{(k)}\}$ exist. Let d^* be an accumulation point such that $|\lambda_n(C + D^*)| > |\lambda_{n+1}(C + D^*)|$, then d^* is the unique global minimizer, and $\{d_i^{(k)}\}$ converges to d^*.*

PROOF: The boundedness of the sequence comes from the monotonic decreasing of function $V(d^{(k)})$. If the boundedness is not true, then there must be a subsequence with index $k \in K_0$ such that $\|d^{(k)}\|_2 \to +\infty$ for $k \in K_0$. Repeating the relevant arguments in Theorem 3.1, we would generate a contradiction to the decreasing property of $V(d^{(k)})$. Since $d^{(k)}$ are bounded sequence, there must be at least one accumulation point.

Let d^* be an accumulation point such that $d^{(k)} \to d^*$ for k in some index set K_0. Under the condition $|\lambda_n(C + D^*)| > |\lambda_{n+1}(C + D^*)|$ we can show that d^* must be a critical point of $V(d)$ such that $\nabla_d V(d^*) = 0$. Suppose that this is not the case. According to the line search and the continuity of $\nabla V(d)$ around d^* we must have

$$\lim_{k \in K_0, k \to \infty} \alpha^{(k)} = 0,$$

and

$$V(d^{(k)} - \hat{\alpha}^{(k)}\nabla V(d^{(k)})) > V(d^{(k)}) - \frac{\hat{\alpha}^{(k)}}{2}\|\nabla V(d^{(k)})\|^2, \qquad (36)$$

here $\hat{\alpha}^{(k)} = 2\alpha^{(k)}$. Note that there is

$$\lim_{\hat{\alpha}^{(k)} \to 0} \frac{V(d^* - \hat{\alpha}^{(k)}\nabla_d V(d^*)) - V(d^*)}{\hat{\alpha}^{(k)}} = -\|\nabla_d V(d^*)\|^2.$$

From the continuity of the second-order derivatives we have, for sufficiently large $k \in K_0$,

$$\frac{V(d^{(k)} - \hat{\alpha}^{(k)} \nabla_d V(d^{(k)})) - V(d^{(k)})}{\hat{\alpha}^{(k)}} = -\|\nabla_d V(d^{(k)})\|^2 + O(\alpha^2) \tag{37}$$
$$\leq -\frac{1}{2}\|\nabla_d V(d^{(k)})\|^2.$$

Clearly, (37) is contradictory to (36). Hence there must be $\nabla_d V(d^*) = 0$. The solution d^* is a global minimizer follows from Theorem 3.2 $\square$

We conclude this section with a remark. The above method can be easily extended to Frobenius norms with "weight". In some applications, the correlation between some forward rates are more important than the rest of the correlations. So we may want to ensure in the calibration process that the importance is properly emphasized. To this purpose we consider the Frobenius norm with "weights":

$$\|A\|_{W,F}^2 = \|\sqrt{W}A\sqrt{W}\|_F^2,$$

with

$$W = diag(w_1, w_2, \ldots, w_n)$$

a diagonal matrix with positive entries. If we think the correlation of the first i_0 forward rates are more important than the correlations between other rates, we can take $w_i = 1$, for $i = 1, \ldots, i_0$, while taking $w_i < 1$, for $i > i_0$. For computations we only need to substitute C in the aforementioned algorithm by $\sqrt{W}C\sqrt{W}$.

3.2. **Eigenvalue problem for the calibration of input prices.** The calibration to input prices has been formulated in the concise form (21). The corresponding Lagrange multiplier problem is

$$\min_d \max_X L(X, d), \tag{38}$$

where

$$L(X, d) = -\left((X - X_0)^T A(X - X_0)\right)^2 + 2\sum_{i=1}^{N_P} d_i(X^T G_i X - h_i). \tag{39}$$

Note that for simplicity we have used $\{G_i, h_i\}$ in place of $\{G_{m,n}, \zeta_{m,n}^2\}$, and have dropped the $\sim$ sign over X_0. To facilitate discussions we again denote the value function for the outer minimization problem by

$$V(d) = \max_X L(X, d), \tag{40}$$

18

and the feasible set of solutions by

$$\mathcal{F} = \{X | X^T G_i X = h_i, i = 1, \ldots, N_P\}.$$

By virtue of the positive-definiteness of the matrix A, the maximizer of $L(X, d)$ is finite for fixed d, and $V(d)$ therefore exists for all d. It is obvious that the Lagrange function is smooth in X. Hence any solutions to the maximization problem (40) must be a critical point of the Lagrange function, satisfying the following first-order condition

$$\left((X - X_0)^T A(X - X_0)\right) A(X - X_0) = (\sum d_i G_i)X$$
$$= (\sum d_i G_i)(X - X_0) + (\sum d_i G_i)X_0. \tag{41}$$

For clarity we denote $B_d = \sum d_i G_i$ and $Y = X - X_0$. Equation (41) then reads

$$[(Y^T AY)A - B_d]Y = B_d X_0. \tag{42}$$

The above equation can be solved through eigenvalue decomposition. Let $(\lambda_i, \mathbf{u}_i)$ be the eigen-pairs of (B_d, A) such that

$$B_d \mathbf{u}_i = \lambda_i A \mathbf{u}_i, \tag{43}$$

$$\mathbf{u}_i^T A \mathbf{u}_i = 1, \quad \mathbf{u}_i^T B_d \mathbf{u}_i = \lambda_i, \tag{44}$$

$$\lambda_i \geq \lambda_{i+1}, \quad i = 1, 2, \ldots, N - 1, \tag{45}$$

and define by U the A-orthogonormal eigenvector matrix

$$U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N). \tag{46}$$

To solve for Y, we let $\alpha = Y^T AY$ and pre-multiply U^T to equation (42), we then arrive at

$$[\alpha I - \Lambda]U^{-1}Y = U^T B_d X_0, \tag{47}$$

where

$$\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$$

is the eigenvalue matrix. From (47) we obtain the solution to (42) as

$$Y \equiv \tilde{Y}(\alpha) = U[\alpha I - \Lambda]^{-1} U^T B_d X_0,$$

where the scalar α is subject to the nonlinear equation

$$\tilde{Y}^T(\alpha) A \tilde{Y}(\alpha) = \alpha,$$

which can be solved with a few steps of iteration. When $X_0 = 0$, in particular, (42) becomes an eigenvalue problem and admits multiple solutions, given by

$$Y_i = \sqrt{\max\{\lambda_i, 0\}} \mathbf{u}_i, \quad i = 1, 2, \ldots, N.$$

19

To fix idea we will from now on concentrate on the special case of $X_0 = 0$. The case $X_0 \neq 0$ is a trivial extension to this special case. When $X_0 = 0$, function $L(X, d)$ achieves it maximum at

$$Y = Y_1, \tag{48}$$

and consequently

$$
\begin{aligned}
V(d) &= \max_{j, \lambda_j \geq 0} -(Y_j^T A Y_j)^2 + 2 \sum_{i=1}^{N} d_i (Y_j^T G_i Y_j - h_i) \\
&= \max_{\lambda_j \geq 0} (\lambda_j^+)^2 - 2 \sum_{i=1}^{N} d_i h_i \\
&= (\lambda_1^+)^2 - 2 \sum_{i=1}^{N} d_i h_i,
\end{aligned}
$$

where $\lambda^+ = max(\lambda, 0)$. It is well-known in matrix theory that the eigenvector corresponding to the largest eigenvalue of a positive semi-definite matrix is the smoothest one amongst all eigenvectors. Very interestingly, the result in (48) establishes the connection between the smoothest fit of volatility surface and the smoothest eigenvector of a generalized eigenvalue problem. Since it usually takes no more than a few steps of iterations to achieve $\lambda_1 > 0$, we assume $\lambda_1 > 0$ in subsequent discussions.

For the analytical properties of $\lambda_1(d), Y(d)(= Y_1(d))$ and $V(d)$ we have

Theorem 3.5. *If $\lambda_1(d) > \lambda_2(d)$, then*

(1) *$(\lambda_1(d), Y(d))$ is differentiable with respect to d locally;*
(2) *$V(d)$ is differentiable in d locally;*
(3) *the gradient of $V(d)$ is*

$$
\nabla_d V(d) = 2 \begin{pmatrix} Y^T G_1 Y - h_1 \\ Y^T G_2 Y - h_2 \\ \vdots \\ Y^T G_{N_P} Y - h_{N_P} \end{pmatrix};
$$

(4) *the elements of the Hessian matrix are given by*

$$H_{ij}(d) \equiv \frac{\partial^2 V}{\partial d_i \partial d_j} = 2 Y^T G_i U \Phi^{-1} U^T G_j Y, \tag{49}$$

where

$$\Phi = \lambda_1 I + 2 \begin{pmatrix} \lambda_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} - \Lambda, \tag{50}$$

20

and the Hessian is a non-negative definite matrix.

PROOF: The proof for statements 1-3 is similar to that of Theorem 3.2, so we only prove the last statement.

The solution Y satisfies equation (42) for $X_0 = 0$. Differentiate both sides of the equation with respect to d_i we have

$$(Y^T AY)A\frac{\partial Y}{\partial d_i} + 2AYY^T A\frac{\partial Y}{\partial d_i} = G_i Y + B_d\frac{\partial Y}{\partial d_i}.$$

From the above equation we can solve for $\frac{\partial Y}{\partial d_i}$:

$$\left[\lambda_1 A + 2AYY^T A - B_d\right]\frac{\partial Y}{\partial d_i} = G_i Y,$$

where we have put $Y^T AY = \lambda_1$. Pre-multiply both sides by U^T we then have

$$\left[\lambda_1 U^T AU + 2U^T AYY^T AU - U^T B_d U\right]U^{-1}\frac{\partial Y}{\partial d_i}$$

$$= \left[\lambda_1 I + 2\begin{pmatrix} \lambda_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} - \Lambda\right]U^{-1}\frac{\partial Y}{\partial d_i}$$

$$= U^T G_i Y,$$

i.e.,

$$\Phi U^{-1}\frac{\partial Y}{\partial d_i} = U^T G_i Y,$$

where matrix Φ is defined in (50). When $\lambda_1 > \lambda_2 \geq \lambda_i, i \geq 3$, Φ is a positive definite matrix, and it follows that

$$\frac{\partial Y}{\partial d_i} = U\Phi^{-1}U^T G_i Y. \tag{51}$$

Component-wise the third statement of the theorem reads

$$\frac{\partial V}{\partial d_i} = 2(Y^T G_i Y - h_i), \qquad i = 1, 2, \dots, N_P.$$

Differentiating this equation with respect to d_j produces the elements of Hessian:

$$\frac{\partial^2 V}{\partial d_i \partial d_j} = 4Y^T G_i\frac{\partial Y}{\partial d_j}$$

$$= 4Y^T G_i U\Phi^{-1}U^T G_j Y.$$

Denote

$$G = [G_1, G_2, \dots, G_{N_P}],$$

21

and define

$$G \otimes Y = [G_1 Y, G_2 Y, \ldots, G_N Y].$$

The non-negative definiteness of the Hessian follows from expression

$$H(d) = 4(G \otimes Y)^T U \Phi^{-1} U^T (G \otimes Y).$$

This completes the proof $\quad \square$

We make the following remarks for future references. For a calibrated model there are

$$0 = \frac{\partial V(d^*)}{\partial d_i} = 2(Y^T G_i Y - h_i), \quad i = 1, 2, \ldots, N.$$

This means that implied Black's volatilities are fit. The above equations allow us to treat inputs h_i's as the functions of d^*:

$$h_i = h_i(d^*) \equiv Y^T G_i Y \Big|_{d=d^*}.$$

Differentiating h_i with respect to d^* we obtain

$$\begin{aligned} \frac{\partial h_i}{\partial d_j^*} &= 2 Y^T G_i \frac{\partial Y}{\partial d_j} \Big|_{d=d^*} \\ &= \frac{1}{2} \frac{\partial^2 V(d^*)}{\partial d_i \partial d_j} = \frac{1}{2} H_{ij}(d^*). \end{aligned} \tag{52}$$

This relation indicates that, for a calibrated model, the Hessian defines the sensitivities of the input prices with respect to the Lagrange multipliers. It makes more sense, however, to know the opposite, that is, the sensitivities of Lagrange multipliers with respect to the input prices. To obtain such sensitivities it is a matter of computing matrix inverse, since (Rockafellar, 1970)

$$\frac{\partial d_j^*}{\partial h_i} = \left(\left(\frac{\partial h_i}{\partial d_j^*} \right)^{-1} \right)_{ij} = 2(H^{-1}(d^*))_{ij}. \tag{53}$$

The above result will be used later to calculate the sensitivities of a derivative instrument with respect to the benchmark instruments.

To ensure that min-max problem has at least one solution, we introduce the concept of *non-arbitrageable implied volatilities*.

Definition 3.1. *We call $h = \{h_i\}$ a set of non-arbitrageable implied volatilities if there is an $\epsilon_h > 0$ such that for any $\{\epsilon_i\}$ with $\epsilon_i \leq \epsilon_h$, there is at least one solution X to*

$$X^T G_i X = h_i + \epsilon_i, \quad i = 1, 2, \ldots, N.$$

We should understand the above concept from the viewpoint of price-volatility corre-spondence. We anticipate that, for a set of realistic prices of market instruments, the market model with a reasonable number of driving factors should be able to "rationalize" the prices through generating a corresponding volatility surface. Furthermore, we want to see that small changes in the prices will be accommodated by the proper variation of the volatility surface. If such accommodation does not happen, then either the model is suggesting the existence of an arbitrage opportunity, or the model simply has no enough dimensions, driving factors in specific, to describe the reality.

For the existence of the global minimizer we have

Theorem 3.6. *If $h = \{h_i\}$ is a set of no-arbitrage implied volatilities, then there is at least one solution to problem $V(d)$. Also, any local minimum is a global minimum.*

PROOF: Given non-empty $\mathcal{F}$, we have

$$V(d) \geq \max_{X \in \mathcal{F}} -((X - X_0)^T A (X - X_0))^2 \equiv V^*$$

for some bounded value V^* due to the positiveness of A. Assume that there exist $d^{(j)} \to \infty$ such that $V(d^{(j)}) \to V^*$ monotonically from above, then, since $\mathcal{F}$ is not empty, we can choose a sequence of $X^{(j)}$ such that

$$(X^{(j)})^T G_i X^{(j)} - h_i = \epsilon^* \text{sign}(d^{(j)}),$$

for some fixed $\epsilon^* > 0$. We then will have

$$V(d^{(j)}) \to +\infty,$$

contradicting to the assumption of monotonic decreasing of $V(d^{(j)})$. Hence, any sequence $\{d^{(j)}\}$ such that $V(d^{(j)}) \to V^*$ monotonically must be bounded and have an accumulation point, and the accumulation point is a solution.

The conclusion that any local minimum is also a global minimum follows from the convexity of $V(d)$ □

For the uniqueness of the solution we have

Theorem 3.7. *Let d^* be a minimizer of $V(d)$. If $\lambda_1(d^*) > \lambda_2(d^*)$ and $H(d^*)$ is positive defi-nite, then d^* is the unique minimizer of $V(d)$ and $Y(d^*)$ solves the constrained minimization problem (21).*

PROOF: If $\lambda_1(d^*) > \lambda_2(d^*)$, then according to Theorem 3.5 $V(d)$ is differentiable near d^* and $\nabla_d V(d^*) = 0$. Moreover, the positive-definiteness of $H(d^*)$ implies that d^* must be

the only local minimizer in its immediate neighborhood. Assume there is another minimizer, say, $d^{**} \neq d^*$, then by the linearity of $L(X, d)$ in d and the convexity of $V(d)$, we have $V(d(t)) = V(d^*), d(t) = td^* + (1-t)d^{**}$ for all $t \in (0, 1)$. That means $V(d^*)$ is not the only minimum in its immediate neighborhood, which is a contradiction.

The conclusion that the solution to the Lagrange multiplier problem solves the constrained minimization problem is obvious (and has been given in Theorem 3.3) □

Similar to the previous section, we can show the convergence of the gradient-based algorithm. Yet, when Hessian has been obtained in closed-form, we should definitely utilize the Hessian and use a Hessian-based algorithm for the numerical solution.

4. Sensitivity with respect to the input prices

In this section we discuss the pricing and hedging of a LIBOR derivative with a model calibrated to the prices of benchmark instruments and correlation matrices. The theory we will develop is in the spirit of Avellaneda *et al.* (1998) for the calibration of equity derivatives via relative-entropy minimization. Assume that the derivative matures at time T_1, and commits a sequence of contingent cash flows $\{F_i\}$, which depend on forward rates $\{f_j\}_{j=1}^N$, on time $T_i, i = 1, 2, \ldots, N$. For convenience we let $T_0 = 0$ be the current time. Let V denotes the value of the contingent claim. The equation governing the price V is

$$
\begin{aligned}
&\frac{\partial V}{\partial t} + \frac{1}{2} \sum s_j(t) s_k(t) \hat{C}_{jk}(t) f_j f_k \frac{\partial^2 V}{\partial f_j \partial f_k} + \sum \mu_j f_j \frac{\partial V}{\partial f_j} - \frac{f_0}{1 + t f_0} V = 0, \\
&V(T_1) = \sum \Pi_{j=1}^{i-1} \left(\frac{1}{1 + \Delta T_j f_j(T_1)} \right) F_i,
\end{aligned}
\tag{54}
$$

where $\mu_j(t) = -\gamma_j(t) \cdot \sigma_{j+1}(t)$ is the drift term of forward rate $f_j(t)$, and f_0 is the term rate for the period $[0, T_1]$. The solution of the equation can be expressed as

$$
V(0, \mathbf{f}(0)) = \frac{1}{1 + f_0 T_1} E^{d^*} \left[\sum \Pi_{j=1}^{i-1} \left(\frac{1}{1 + f_j \Delta T_j} \right) F_i \right],
\tag{55}
$$

where $\mathbf{f}(t) = \{f_j(t)\}$ and d^* stands for the calibrated measure. The expectation in (55) can be calculated by Monte-Carlo simulations or finite difference methods.

Now we consider hedging the short position with the benchmark instruments, for which we need to calculate the *hedge ratios*. Let the prices of the benchmarks be $\{C_j\}$. By the chain rule we have

$$
\frac{\partial V}{\partial C_j} = \sum_{i=1}^N \frac{\partial V}{\partial d_i} \frac{\partial d_i}{\partial C_j} = \sum_{i=1}^N \frac{\partial V}{\partial d_i} \frac{\partial d_i}{\partial h_j} \frac{\partial h_j}{\partial C_j},
\tag{56}
$$

and we know that one must calculate three derivative chains

$$\left\{\frac{\partial h_j}{\partial C_j}\right\}, \quad \left\{\frac{\partial d_i}{\partial h_j}\right\} \quad \text{and} \quad \left\{\frac{\partial V}{\partial d_i}\right\}.$$

The first two derivative chains are easy to obtain. Recall (53), we have

$$\frac{\partial d_i}{\partial h_j} = 2(H^{-1})_{i,j},$$

where H is the Hessian matrix whose elements are given by (49). The partial derivatives of volatilities with respect to prices can be calculated from the Black's formula (which includes a caplet as a special case):

$$\frac{\partial h_j}{\partial C_j} = 1 \left/ \frac{\partial C_j}{\partial h_j} \right. = \frac{1}{B_{m,n}^S[R_{m,n}n(g_+)g'_+(h_j) - Kn(g_-)g'_-(h_j)]},$$

where $g_\pm$ are given in (7), and

$$n(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

$$g'_\pm(h) = -\frac{1}{2}h^{-\frac{3}{2}}\ln\frac{R_{m,n}}{K} \pm \frac{1}{4}h^{-\frac{1}{2}}.$$

Finally, to calculate $\frac{\partial V}{\partial d_i}$ we differentiate equation (54) with respect to d_i:

$$\frac{\partial V_i}{\partial t} + \frac{1}{2}\sum s_j(t)s_k(t)\hat{C}_{jk}(t)f_jf_k\frac{\partial^2 V_i}{\partial f_j\partial f_k} + \sum \mu_jf_j\frac{\partial V_i}{\partial f_j} - \frac{f_0}{1+tf_0}V_i$$

$$= -\frac{1}{2}\sum \left(\frac{\partial s_j(t)}{\partial d_i}s_k(t) + s_j(t)\frac{\partial s_k(t)}{\partial d_i}\right)\hat{C}_{jk}(t)\frac{\partial^2 V}{\partial f_j\partial f_k}, \quad (57)$$

where $V_i \equiv \frac{\partial V}{\partial d_i}$. As given by (51),

$$\frac{\partial s_j(t)}{\partial d_i} = \left(\frac{\partial Y}{\partial d_i}\right)_j = \left(U\Phi^{-1}U^TG_iY\right)_j.$$

Functions $\{V_i\}$ can be solved backwardly in time together with V. From the equation we know that we can write

$$\frac{\partial V}{\partial d_i} = E^{d^*}\left[-\frac{1}{2}\int_0^{T_1}\frac{1}{1+tf_0}\sum\left(\frac{\partial s_j(t)}{\partial d_i}s_k(t) + s_j(t)\frac{\partial s_k(t)}{\partial d_i}\right)\hat{C}_{jk}(t)f_jf_k\frac{\partial^2 V}{\partial f_j\partial f_k}dt\right]$$

Put all calculated partial derivatives back in (56) we obtain the hedging ratio with respect to the ith benchmark instrument.

5. Numerical results

In this section we will see the performance of our calibration methods with two examples. We will comment the accuracy and efficiency of the calibration methods in the due course.

The first is a hypothetical example taken from Rebonato (1999), where the low-rank approximation of a hypothetical correlation matrix is calculated. In this example, we consider a collection of twelve 12-month forward rates with (market) correlation matrix C given by

$$c_{ij}^{market} = LongCorr + (1 - LongCorr) \exp[\beta|t_i - t_j|],$$
$$\beta = d_1 - d_2 \max(t_i, t_j).$$

Here, we take $LongCorr = 0.3, d_1 = -0.12, d_2 = 0.005$, and the reset dates as $t_i = i \times (\frac{1}{2})$. This matrix is of full rank. We use our algorithm to calculate its approximations for various ranks. We feel that it is not necessary to compare the performance of our method with that of Rebonato's (1999), which is attached in the appendix for readers' reference.

The results of rank-3 approximation are given in Table 1 and Figure 1-4. Figure 1 shows the market correlation surface, while Figure 2 shows the model correlation surface. Apparently, the latter is smooth and in good agreement with the market correlation, except the extent of convexity near the diagonal. The effect of "smoothing out" along the diagonal, however, is bound to happen and can not be improved, as the best low-rank approximation can only be obtained by using vectors very close to the leading principal components of the market correlation matrix, which are always the smoothest ones amongst all principal components. Table 1 and Figure 3 display (the difference of) principal components. It can be seen that, while the first principal component of the model is very close to that of the market, there exists visible differences in the second and the third principal components between the two correlation matrices. This is actually a desirable feature because the first component describes the proportional parallel shift of the rates, and it is the most important component that characterizes the correlation between the different rates. The market and model correlation between (a) the first, (b) third, (c) the sixth, and (c) the tenth forward rate and the rest of the forward rates are displayed in Figure 4. It takes less then seven functional evaluations of $V(d)$ to obtain the model correlation.

The second is a practical example taken from Brace *et al.* (1997). In this example, we calibrate the market model for the Sterling Pound to the prices of a set of benchmark instruments and the correlation matrix of Sterling Pound forward rates. Note that from historical data we can only estimate the *spot* correlations (the correlations seen at the moment), but not the *forward* correlations (the correlation seen at a future date). It is very much a convention that we define the *forward* correlations by *time homogeneity*, namely, the correlation between any two forward rates depends on the difference of their reset times

only. This convention allows us to define the market correlations of all forward rates. The market prices of a set of caps and swaptions are listed in Table 2, and the correlation matrix is given in Table 3 (which was calculated with the one-year data of 1994). The first row and column of the table show the maturities of the forward rates. In our calculations, the inputs of prices are taken in the form of implied volatilities of caplets and swaptions. Hence, a pre-processing procedure is adopted to 1) convert the cap prices into corresponding caplet prices by the method of *bootstrapping*, and 2) map the caplet and swaption prices into their implied volatilities, using a root-finding algorithm. For bootstrapping the caplet prices, we need the spot yield curve or zero-coupon bond prices of maturities covering all forward periods of involving forward rates. The zero-coupon bond prices are listed in Table 4, for maturities up to 11 years. The caplet prices and their implied Black volatilities, together with those of swaptions, are listed in Table 5.

First let us take a look at the low-rank approximations to the market correlation matrix. We have calculated the rank- one, two, three, six and ten approximations and the outcomes are visualized in Figure 5-10, where the first figure is the original market correlation surface, and the flat surface plot for rank-one approximation is due to the identical diagonal entries. By vision we shall agree that the approximation improves with increasing rank. Numerically the trend of convergence with respect to the increasing rank is given in Figure 11. For each number of factors the calibration requires less than seven functional valuation (of $V(d)$). In fact this is a very small scale problem for the method developed in this paper.

With calculated low-rank approximations of the correlation matrix, we proceed to compute the forward rate volatilities, $\{s_j^i\}$, from the input Black volatilities of caplets and swaptions. The implementation is made with a Hessian-based unconstrained minimization function in MATLAB ("fminunc" in specific). The results are plot as volatility surfaces from Figure 12 to 16. The magnitude of calibration error relative to the (squares of) input volatilities is of order

$$\frac{\|Y^T G_{m,n} Y - \zeta_{m,n}^2\|_2}{\|\zeta_{m,n}^2\|_2} = O(10^{-4}).$$

For the Black volatility under 20%, this corresponds to less than 0.2% of error (in terms of volatility), which is much less than the usual 1% bid/ask spread in the markets. In calculating the volatility surfaces, we have taken all implied volatilities except those of the last two swaptions. That is, we have altogether 15 inputs. If we add in the second last swaption volatility, the calibration error will rise to 4%. After adding in both of the last

27

two swaption volatilities, the iteration of outer minimization does not converge for even the ten-factor model. The cause is not yet fully understood but we think there probably exists price inconsistence amongst the inputs. We can see that the volatility surfaces look incredibly close yet, as Figure 17 shows, not identical. The closeness can be explained by the facts that caplet prices do not depend on correlation, and the swaption prices in our data set depend only on elements near the diagonal of the correlation matrix, which are close across correlation matrices of various ranks. A positive implication of the closeness is that the calculated forward rate volatilities is not sensitive to the changes in correlation matrix. Again each calibration to prices takes about seven functional valuations. The entire calibration (to both correlations and implied Black's volatilities) is finished within twenty seconds. Note that the gradient-based minimization with "fminunc" takes much longer. Hence the Hessian-based algorithm is highly recommended.

6. CONCLUSION

We believe that we have offered the optimal methodology to calibrate the market model to ATM cap/floor and swaption prices as well as exogenously given correlation matrices. The calibration of the prices and correlation matrix is decoupled into two subproblems. Both of the subproblems are formulated as minimization problems with convex objective functions, and the functional evaluation is achieved by the eigenvalue decomposition of an matrix. The numerical method is very efficient, and robust to the number of driving factors of the model and the number of input prices.

There are other outstanding calibration problems. Calibration to volatility smile is a more challenging problem, and will most likely involve the lognormal model with stochastic volatilities, or a jump-diffusion model. Our methodology does not seem to extend trivially to these extended models. Extracting the "implied correlations" is another interesting issue for the market model. These issues have gone beyond the scope of the current paper and they are left for future studies.

References

[1] Andersen, L. and J. Andreasen (2000). Volatility skews and extensions of the LIBOR market model. *Applied Mathematical Finance*, Vol. 7, No 1, 1-32.

[2] Andersen, L. (1999). A simple approach to the pricing of Bermudan swaptions in the multi-factor LIBOR market model. *Journal of Computational Finance*, Vol. 2, No 3, 5-32.

[3] Avellaneda, M., C. Friedman, R. Holmes and D. Samperi (1998). Calibrating Volatility Surfaces via Relative-Entropy Minimization. Working paper, New York University.

[4] Black, F. (1976). The pricing of commodity contract. *Journal of Financial Economics*, 3, 167-179.

[5] Brace, A., D. Gatarek and M. Musiela (1997). The Market model of interest rate dynamics. *Mathematical Finance*, 7(2), 127-155.

[6] Brotherton-Ratcliffe, R. (1997). The BGM model for Path-dependent swaps. Working paper, General Re Financial Products.

[7] Glasserman, P. and Z. Zhao (1998). Arbitrage-free discretization of lognormal forward LIBOR and swap rate model. Working paper, Columbia University.

[8] 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.

[9] Jamshidian, F., (1997). LIBOR and swap market models and measures. *Finance and Stochastic*, 1, 293-330.

[10] Miltersen, K., K. Sandmann, and D. Sondermann (1997). Closed-form solutions for term structure derivatives with lognormal interest rates. *Journal of Finance*, 409-430.

[11] Pedersen, M.B. (1999). On the LIBOR market models. Working paper, SimCorp Financial Research.

[12] Rebonato, R. (1999). Calibrating the BGM model. *RISK*, March, 74-79.

[13] Rockafellar, R.T. (1970). Convex Analysis. Princeton, New Jersey: Princeton University Press.

[14] Sidennius, J., and P. Soegaard-Andersen (1999) The valuation of constant maturity swaps, caps and floors. Working paper, SimCorp Financial Research.

[15] Sidennius, J. (2000). LIBOR market models in practice. *Journal of Computational Finance*, 3(3), 5-26.

[16] Stewart, G.W. and J.-G. Sun (1990). Matrix Perturbation Theory. Academic Press.

[17] Zhang, Z. and L. Wu (2001). Optimal low-rank approximation to correlation matrix. Submitted the *Linear Algebra and Its Applications*.

Appendix A. Constrained Minimization by Parameterization

To solve for (22), Rebonato (1999) considers solution of the form

$$X = BB^T, \tag{A.1}$$

where B is an N by n matrix whose elements are of the parametric functions

$$b_{jk} = \cos(\theta_{jk})\Pi_{l=1}^{k-1}\sin(\theta_{jl}), \quad k = 1, \dots, n-1,$$

$$b_{jn} = \Pi_{l=1}^{n-1}\sin(\theta_{jl}). \tag{A.2}$$

Note that representation (A.1) guarantees the rank of X to be less or equal to n, while the parameterization (A.2) ensures the "one-diagonal" condition as we have

$$\sum_{k=1}^{n} b_{jk}^2 = 1, \qquad j = 1, 2, \dots, N,$$

for any angles $\{\theta_{jk}\}$. Given the representation and parameterization which effectively remove the constraint, Rebonato proceeds to solve the unconstrained problem

$$\min_{\{\theta_{jk}\}} \|C - B(\{\theta_{jk}\})B^T(\{\theta_{jk}\})\|_F,$$

with standard unconstrained minimization methodologies. This is a nonlinear optimization problem with $N \times n$ unknowns. In financial applications, this number can go as high as $80 \times 4 = 320$, which then poses a horrendous challenge to any existing methodologies.
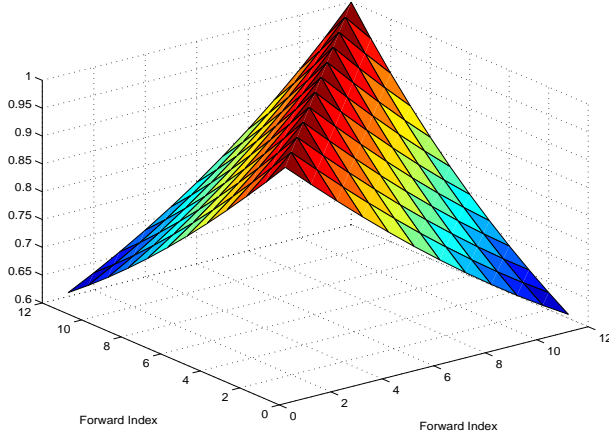
FIGURE 1. Market correlation surface
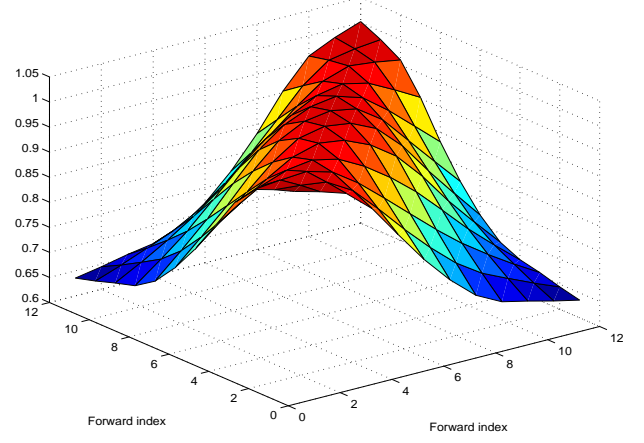
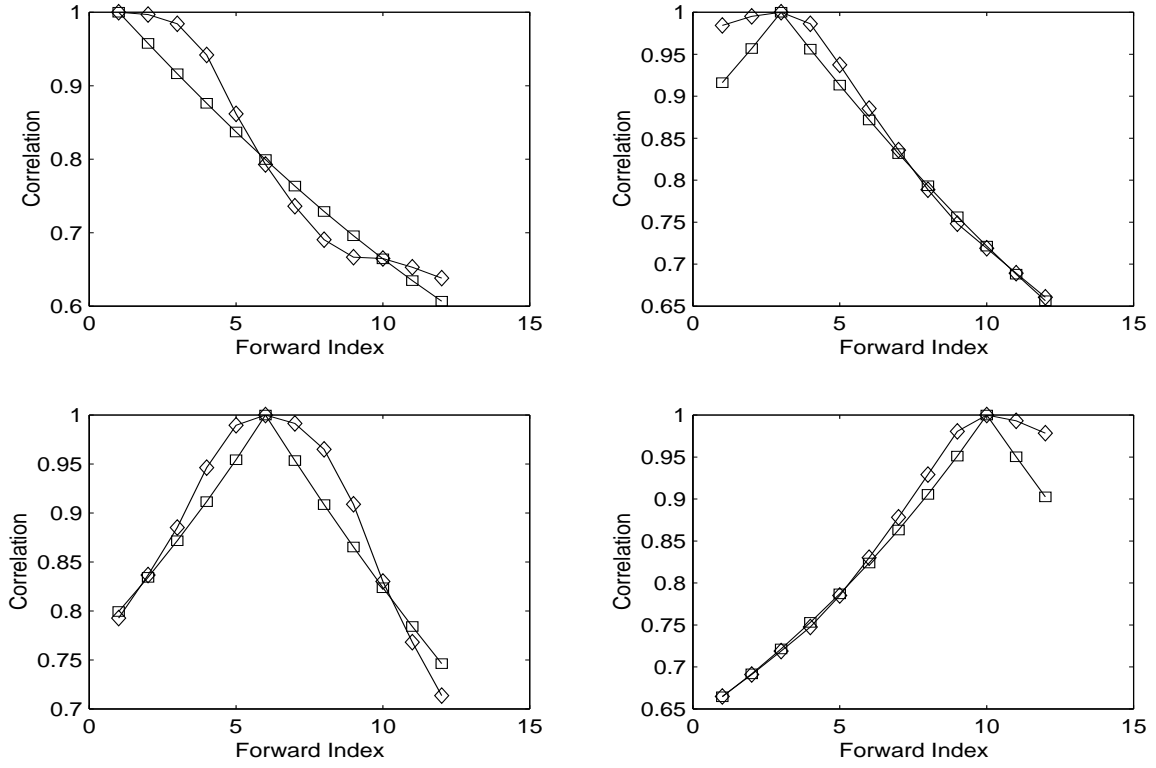FIGURE 2. Rank 3 model correlation surface

FIGURE 3. The market and model correlations between (a) the first, (b) the third, (c) the sixth and (d) the tenth forward rates and the rest of the forward rates obtained using three-factor iterative model.

TABLE 1. PRINCIPAL COMPONENTS OF THE RANK-ONE CORRECTION

U_1	U_2	U_3	$U_{a,1}$	$U_{a,2}$	$U_{a,3}$
0.86	-0.39	0.22	0.87	-0.42	0.27
0.89	-0.38	0.17	0.90	-0.40	0.20
0.92	-0.34	0.08	0.92	-0.37	0.11
0.93	-0.27	-0.04	0.95	-0.31	-0.04
0.95	-0.18	-0.15	0.96	-0.20	-0.20
0.95	-0.07	-0.22	0.96	-0.07	-0.26
0.95	0.05	-0.23	0.96	0.06	-0.27
0.95	0.17	-0.17	0.96	0.19	-0.22
0.93	0.27	-0.07	0.95	0.31	-0.08
0.91	0.35	0.06	0.92	0.38	0.10
0.88	0.40	0.18	0.89	0.41	0.21
0.85	0.41	0.24	0.85	0.43	0.29
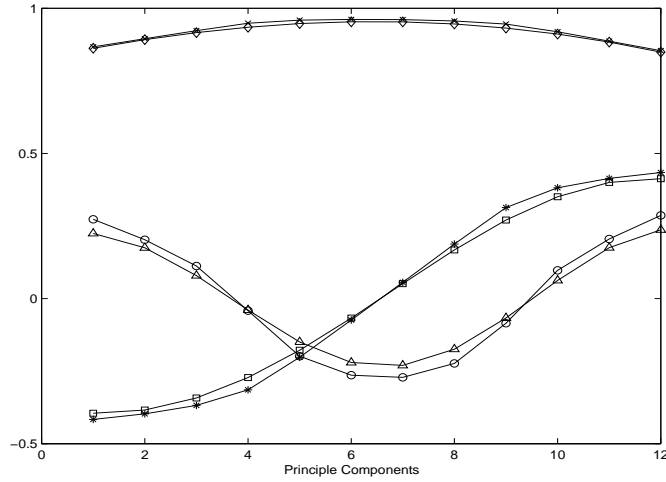
FIGURE 4. The first three principal components of the market and model correlation matrixes

TABLE 2. INPUT PRICES OF CAPS AND SWAPTIONS (FEB. 3, 1995)

Contract type	Length	A-T-M strike(%)	Black vol. (%)	Market price(bp)
cap	1	7.88	15.5	27
cap	2	8.39	17.75	100
cap	3	8.64	18	185
cap	4	8.69	17.75	267
cap	5	8.79	17.75	360
cap	7	8.9	16.5	511
cap	10	8.89	15.5	703
	Option maturity ×Swap length			
Swaption	0.25× 2	8.57	16.75	50
Swaption	0.25× 3	8.75	16.5	73
Swaption	1× 4	9.1	15.5	172
Swaption	0.25× 5	8.9	15	103
Swaption	0.25× 7	9	13.75	123
Swaption	0.25× 10	8.99	13.25	151
Swaption	1× 9	9.12	13.25	271
Swaption	2× 8	9.16	12.75	312

TABLE 3. HISTORICAL CORRELATION MATRIX FOR THE GBP FORWARD RATES

	0.25	0.5	1	1.5	2	2.5	3	4	5	7	9
0.25	1.0000	0.8415	0.6246	0.6231	0.5330	0.4287	0.3274	0.4463	0.2439	0.3326	0.2625
0.5	0.8415	1.0000	0.7903	0.7844	0.7320	0.6346	0.4521	0.5812	0.3439	0.4533	0.3661
1	0.6246	0.7903	1.0000	0.9967	0.8108	0.7239	0.5429	0.6121	0.4426	0.5189	0.4251
1.5	0.6231	0.7844	0.9967	1.0000	0.8149	0.7286	0.5384	0.6169	0.4464	0.5233	0.4299
2	0.5330	0.7320	0.8108	0.8149	1.0000	0.9756	0.5676	0.6860	0.4969	0.5734	0.4771
2.5	0.4287	0.6346	0.7239	0.7286	0.9756	1.0000	0.5457	0.6583	0.4921	0.5510	0.4581
3	0.3274	0.4521	0.5429	0.5384	0.5676	0.5457	1.0000	0.5942	0.6078	0.6751	0.6017
4	0.4463	0.5812	0.6121	0.6169	0.6860	0.6583	0.5942	1.0000	0.4845	0.6452	0.5673
5	0.2439	0.3439	0.4426	0.4464	0.4969	0.4921	0.6078	0.4845	1.0000	0.6015	0.5200
7	0.3326	0.4533	0.5189	0.5233	0.5734	0.5510	0.6751	0.6452	0.6015	1.0000	0.9889
9	0.2625	0.3661	0.4251	0.4299	0.4771	0.4581	0.6017	0.5673	0.5200	0.9889	1.0000

Forward rates were assumed constant on the intervals between the terms.

TABLE 4. PRICES OF ZERO-COUPON BONDS

Tenor	Maturity
0.00	1.00000000
0.25	0.98317518
0.50	0.96533801
1.00	0.92713249
1.50	0.88814477
2.00	0.84964678
2.50	0.81226987
3.00	0.77629645
4.00	0.71122696
5.00	0.64912053
7.00	0.54020582
9.00	0.45339458
10.00	0.41531609
11.00	0.37873810

TABLE 5. STRIPPED CAPLET PRICES AND SWAPTION PRICES

Contract type	Option maturity ×Swap length	ATM Strike (%)	Black Vol. (%)	Market Price(bp)
Caplet	0.25× 0.25	7.88	0.15	1.59
Caplet	0.5× 0.25	7.88	0.15	13.20
Caplet	1× 0.25	8.39	0.19	19.26
Caplet	2× 0.25	8.64	0.18	25.36
Caplet	3× 0.25	8.69	0.17	24.61
Caplet	4× 0.25	8.79	0.18	29.92
Caplet	5× 0.25	8.90	0.14	29.20
Caplet	7× 0.25	8.89	0.13	21.88
Caplet	9× 0.25	8.89	0.13	17.40
Swaption	0.25×2	8.59	0.16	50.00
Swaption	0.25×3	8.79	0.16	73.00
Swaption	1×4	0.0910	0.16	172.00
Swaption	0.25×5	8.95	0.15	103.00
Swaption	0.25×7	9.04	0.14	123.00
Swaption	0.25×10	9.02	0.13	151.00
Swaption	1×9	9.14	0.13	271.00
Swaption	2×8	9.18	0.13	312.00
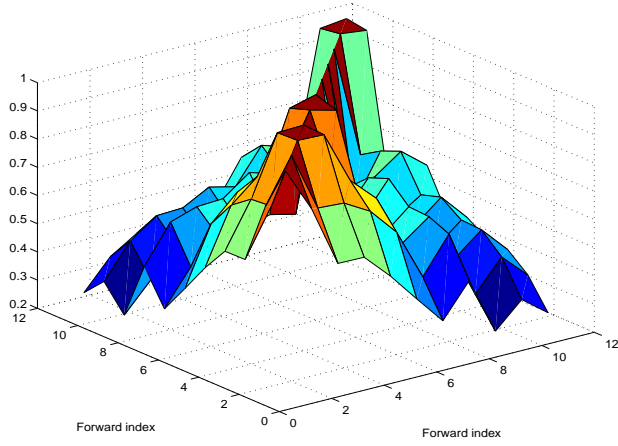
FIGURE 5. Market correlation surface
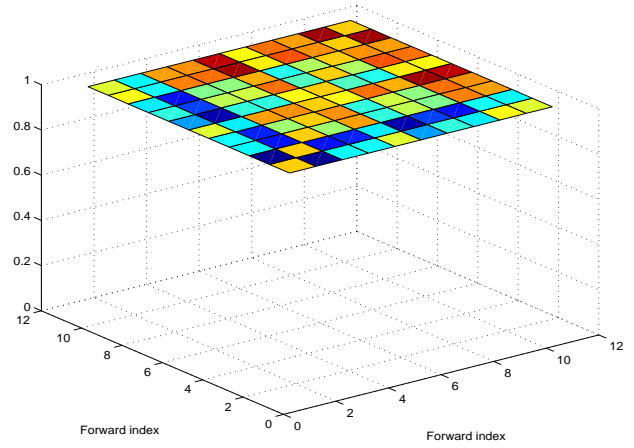
FIGURE 6. Correlation surface of one-factor model
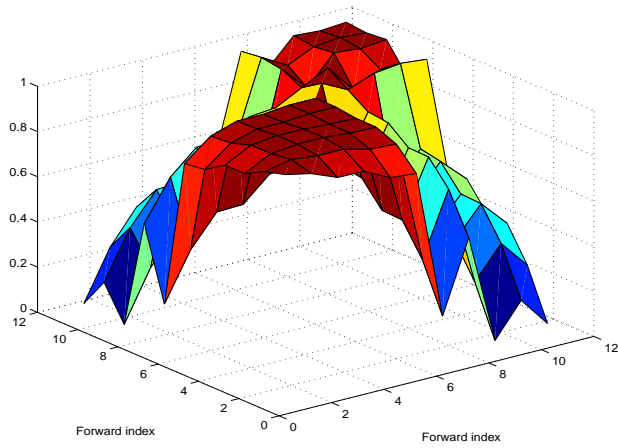
FIGURE 7. Correlation surface of two-factor model
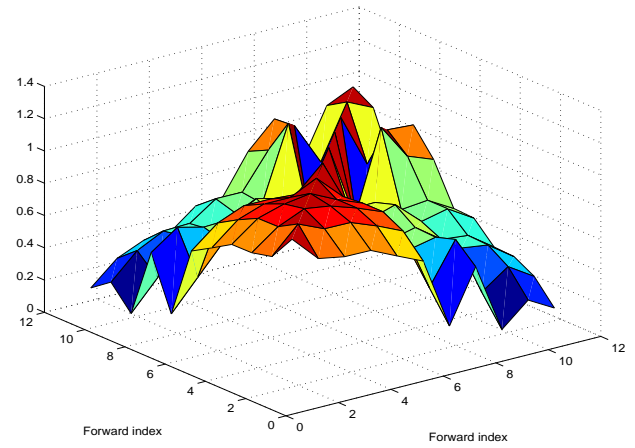
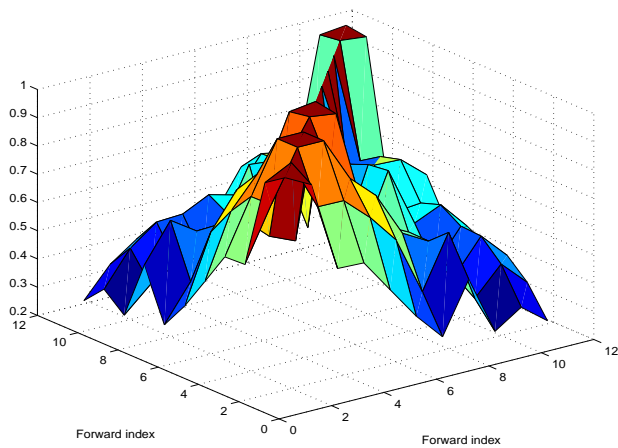FIGURE 8. Correlation surface of three-factor model

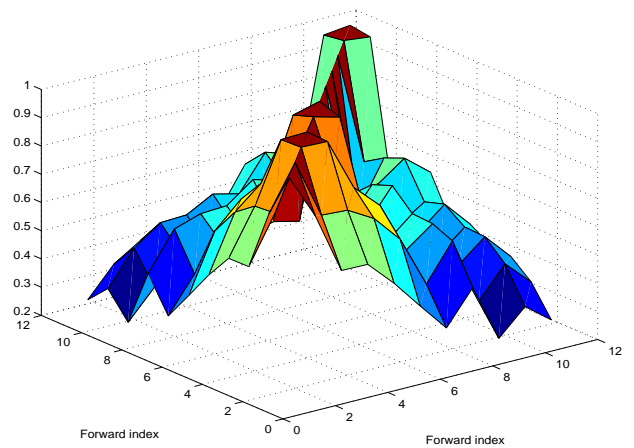FIGURE 9. Correlation surface of six-factor model
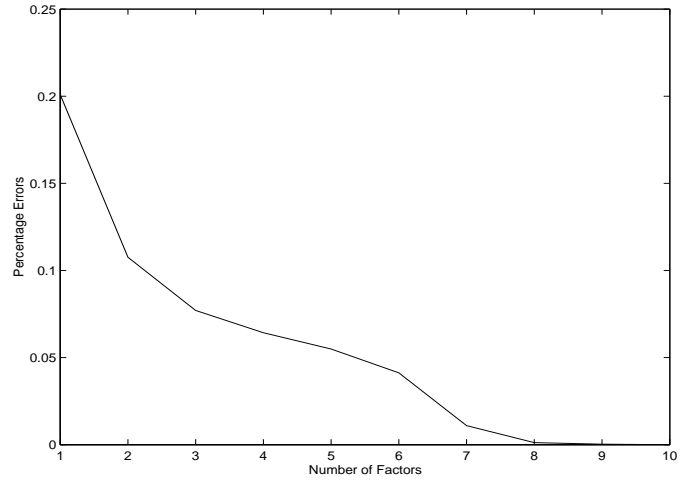

FIGURE 10. Correlation surface of ten-factor model

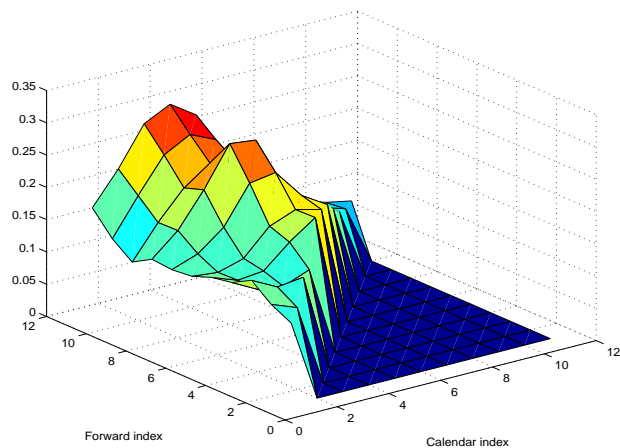FIGURE 11. Trend of Convergence with increasing Rank

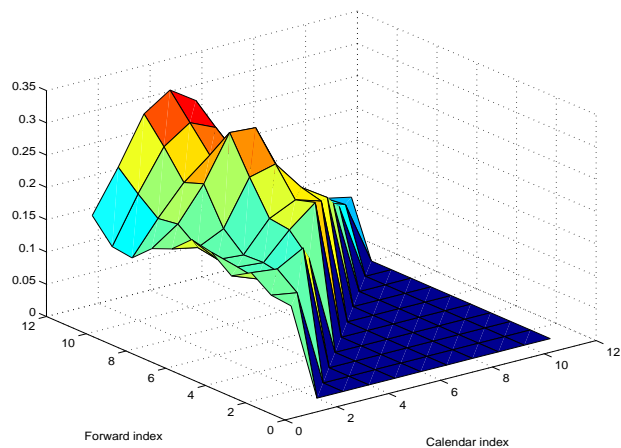FIGURE 12. Volatility surface of one-factor model
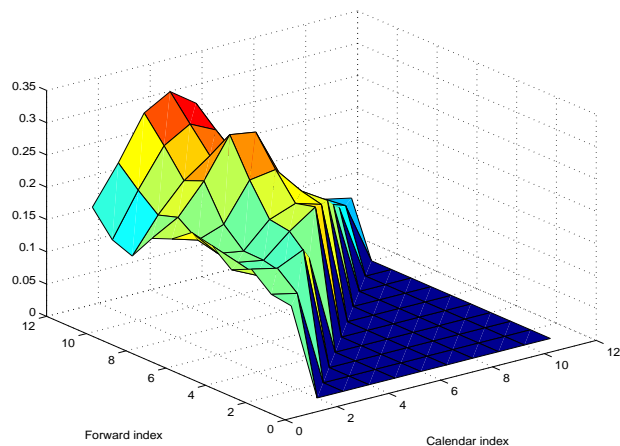

FIGURE 13. Volatility surface of two-factor model

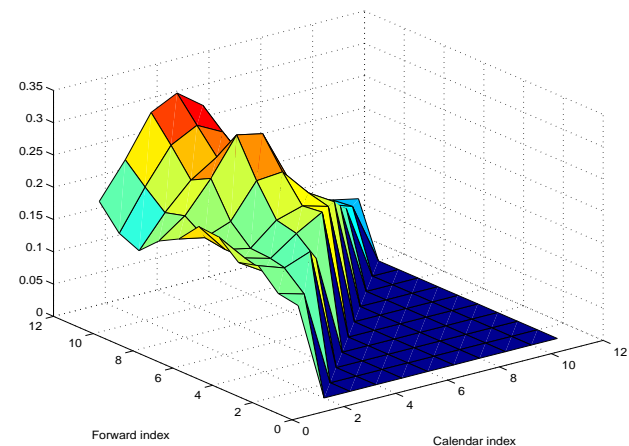FIGURE 14. Volatility surface of three-factor model

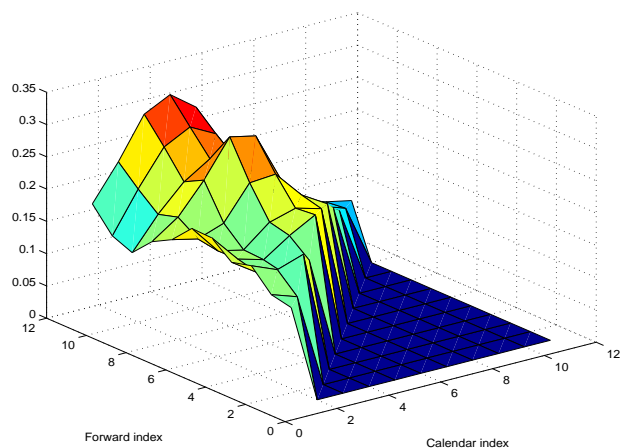FIGURE 15. Volatility surface of six-factor model
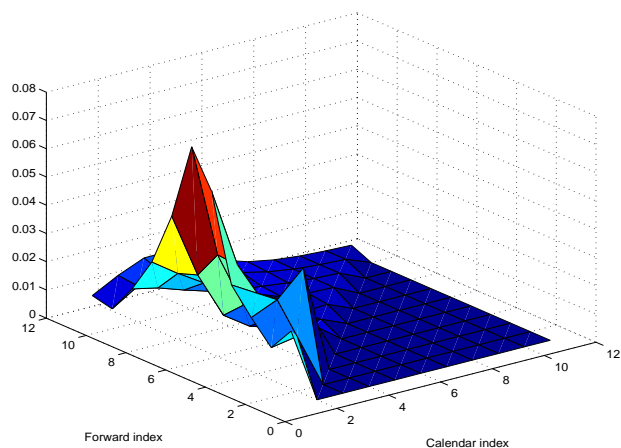

FIGURE 16. Volatility surface of ten-factor model

FIGURE 17. Difference between one- and ten-factor models