On the distributional distance between the Libor and the Swap market models *

Damiano Brigo
Product and Business Development Group
Banca IMI, SanPaolo IMI Group
Corso Matteotti 6, 20121 Milano, Italy
Fax: +39 02 7601 9324
E-mail: brigo@bancaimi.it
Web Site: http://www.damianobrigo.it

Jan Liinev†
Department of Applied Mathematics and Computer Science
University of Ghent
Krijgslaan 281/S9 B-9000 GENT BELGIUM
Tel: +32 (0)9 2644884
E-mail: Jan.Liinev@rug.ac.be

First version: June 1, 2001. This version: June 5, 2003

Abstract

In this paper we are concerned with the distributional difference of forward swap rates between the lognormal forward–Libor model (LFM) or "Libor market model" and the lognormal forward-swap model (LSM) or "swap market model", the two fundamental models for interest-rate derivatives. To measure this distributional difference, we resort to a "metric" in the space of distributions, the well known Kullback-Leibler information

*Preliminary version. We are grateful towards Lorenzo Finesso for introducing us to the literature on entropy estimators. We are also grateful towards Francesco Rapisarda for suggestions on improvements of Monte Carlo simulations and to Fabio Mercurio for general discussions.

†Jan Liinev is supported by the BOF-project 001104599, whose financial support is gratefully acknowledged.

(KLI). We explain how the KLI can be used to measure the distance of a given distribution from the lognormal (exponential) family of densities, and then apply this to our models' comparison. The volatility of (i.e. standard deviation associated with) the projection of the LFM swap-rate distribution onto the lognormal family is compared to a industry synthetic swap volatility approximation obtained via "freezing the drift" techniques in the LFM. Finally, for some instantaneous covariance parameterizations of the LFM we analyze how the above distance changes according to the parameter values and to the parameterizations themselves, in an attempt to characterize the situations where LFM and LSM are really distributionally close, as is assumed by the market.

1 Introduction

In this paper we are concerned with the difference between two of the most popular and promising interest rate models: the two main market–models. The importance of such models is due to their agreement with well-established market formulae for two basic derivative products. Indeed, the lognormal forward–Libor model (LFM) or "Libor market model" (Miltersen, Sandman and Sondermann (1997), and Brace, Gatarek and Musiela (1997)), prices caps with Black's cap formula, which is the standard formula employed in the cap market. Moreover, the lognormal forward-swap model (LSM) or "swap market model" (Jamshidian (1997)), prices swaptions with Black's swaption formula, which again is the standard formula employed in the swaption market. Since the cap and swaption markets are the two main markets in the interest-rate world, it is important for a model to be compatible with such markets formulae. Recently, many works have investigated the LIBOR market model properties, calibration, approximated lognormality under different measures, and other issues, see for example Brigo and Mercurio (2001), Matsumoto (2001), Rebonato (2003), Sidenius (2000).

Before market models were introduced, Black's formula for either caps or swaptions was based on mimicking the Black and Scholes model for stock options under some simplifying and inexact assumptions on interest rates distributions. The introduction of market models provided a new and rigorous derivation of Black's formulae. However, even with full rigor given separately to the caps and swaptions classic formulae, the LFM and LSM are not compatible. Roughly speaking, if forward Libor rates are lognormal each under its measure, as assumed by the LFM, forward swap rates cannot be lognormal at the same time under the relevant swap measure, as assumed by the LSM instead. There exists a number of empirical works in the literature showing that forward swap rates obtained by lognormal forward Libor rates are not far from being lognormal themselves under the appropriate measures. We refer to the paper of Brace, Dun and Barton (1998) for such analysis, where "closeness" of the two models is analysed based on standard statistical techniques for assessing the deviation from normality of a certain set of data. Moreover, in Chapter 8 of Brigo and Mercurio (2001) some problems related to this matter have been investigated numerically, showing that lognormality almost holds for the swap rate in the LIBOR model.

Here we propose instead the use of a "metric" in the space of distributions as a measure of the difference between the two models. We consider the forward rate dynamics under the appropriate forward measures for the LFM. As we said before, we know that in this setup caps are priced in agreement with Black's cap market formula, provided one chooses a geometric Brownian martingale dynamics for the relevant forward rates, and cap pricing is almost automatic, the degree of automatism depending on the particular instantaneous volatility parameterization chosen in the forward rates dynamics. A fundamental part of the LFM specification which has a relevant impact on the forward swap rates coming from this model is the choice of the instantaneous *covariance* structure in the forward rates dynamics, i.e. of instantaneous volatility *and correlation*. We illustrate some possible parametric choices for both of these structures. We then derive the LFM forward rates dynamics under the swap measure. We simulate such a forward rate dynamics under the swap measure, and compute the resulting swap rate as an algebraic expression in the forward rates thus generated. In this way we simulate the swap rate under the swap measure *in the LFM*. Similarly, we introduce the LSM and show how swaptions are priced in agreement with Black's swaptions formula under the swap measure, implying a lognormal distribution for the forward swap rate, and point out the theoretical incompatibility between the LFM and LSM. We take the distance of the simulated LFM swap rate distribution from the family of lognormal distributions, where the LSM swap rate distribution lies, as a measure of the distributional distance between the LFM and LSM.

We also report an approximated formula which allows to deduce LSM instantaneous volatilities from LFM instantaneous covariances. This formula is rather handy and popular among practitioners, and is based on ignoring drifts and on freezing coefficients in suitable stochastic differential equations for the forward swap rates.

We then introduce the Kullback–Leibler information (KLI) and explain how it can be used to measure the distance of a given distribution from an exponential family. We plan to use this distance to understand how closeness between LFM and LSM changes according to the particular volatility and correlation parameterizations which are chosen for the LFM instantaneous covariance. For each parameterization considered, we also analyze how the distance changes with volatility parameters, correlation parameters, time to maturity, and tenor (length of the underlying swap). Finally, in all these different situations, we compare the approximated formula for swap volatilities based on "ignoring the drift" and "freezing the coefficients" to a volatility based on the KLI. This KLI volatility is obtained via the lognormal distribution which is closest (in the KLI sense) to the true swap–rate distribution of the LFM.

In the paper we report the first results we obtained in this investigation.

2 The Libor market model (LFM)

Let $t = 0$ be the current time. Consider a set $\mathcal{E} = \{T_0, \ldots, T_M\}$ of adjacent expiry-maturity pairs of dates for a family of spanning forward rates. We shall denote by $\{\tau_0, \ldots, \tau_M\}$ the corresponding year fractions, meaning that τ_i is the year fraction associated with the expiry-maturity pair T_{i-1}, T_i for $i > 0$, and τ_0 is the year fraction from settlement to T_0.

Consider the generic forward (LIBOR) rate $F_k(t) := F(t; T_{k-1}, T_k)$, which is "alive" up to time T_{k-1}, where it coincides with the spot (LIBOR) rate $F_k(T_{k-1}) := L(T_{k-1}, T_k)$, where in general $L(S,T)$ is the spot (LIBOR) rate prevailing at time S for the maturity T.

Consider now the probability measure Q^k associated to the numeraire $P(\cdot, T_k)$, i.e. to the price of the zero-coupon bond whose maturity coincides with the maturity of the forward rate. Q^k is often called the *forward adjusted measure for maturity* T_k. Since

$$F_k(t)P(t,T_k) = (P(t,T_{k-1}) - P(t,T_k))/\tau_k.$$

is the price of a tradable asset, F_k follows a martingale under Q^k. The lognormal forward-Libor model (LFM) assumes the following driftless geometric Brownian dynamics for F_k under Q^k:

$$dF_k(t) = \sigma_k(t)\ F_k(t) dZ_k^k(t), \quad t \le T_{k-1}, \tag{1}$$

where $Z_k^k(t)$ is a standard Brownian motion under Q^k, the upper index denoting the measure, and $\sigma_k(t)$ is a deterministic function representing the instantaneous volatility at time t for the forward LIBOR rate F_k. In case we write F_k's dynamics under a measure Q^i with $i \neq k$, F_k does not follow a martingale under Q^i and a drift term appears, yielding a SDE with unknown distributional properties. For the details see for example Chapter 6 of Brigo and Mercurio (2001).

We will often consider piecewise constant instantaneous volatilities,

$$\sigma_k(t) = \sigma_{k,\beta(t)}, \quad \text{where} \quad \beta(t) = m \ \text{ if } \ t \in (T_{m-2}, T_{m-1}].$$

The noises in the dynamics of different forward rates are assumed to be instantaneously correlated according to

$$dZ_i(t)\ dZ_j(t) = d\langle Z_i, Z_j \rangle_t = \rho_{i,j}\ dt.$$

Clearly, lower indices denote components. Upper indices are omitted when the measure is clear from the context or is irrelevant.

A few general remarks are now in order. First, as we said before, we will often assume that the forward rate $F_k(t)$ has instantaneous volatility that is piecewise constant: In particular, the instantaneous volatility of $F_k(t)$ is constant in each "expiry-maturity" time interval (associated to any other forward rate) $T_{m-2} < t \le T_{m-1}$ where it is "alive".

Under this assumption, it is possible to organize instantaneous volatilities in a matrix as follows:

TABLE 1

Instant. Vols	Time: $t \in (0, T_0]$	$(T_0, T_1]$	$(T_1, T_2]$	...	$(T_{M-2}, T_{M-1}]$
Fwd Rate:$F_1(t)$	$\sigma_{1,1}$	Dead	Dead	...	Dead
$F_2(t)$	$\sigma_{2,1}$	$\sigma_{2,2}$	Dead	...	Dead
$\vdots$	...	...	...	...	...
$F_M(t)$	$\sigma_{M,1}$	$\sigma_{M,2}$	$\sigma_{M,3}$	...	$\sigma_{M,M}$

A possibly good structure is the following separable piecewise constant (SPC) formulation:

$$\sigma_k(t) = \sigma_{k,\beta(t)} := \Phi_k \psi_{k-\beta(t)-1} \tag{2}$$

for all t. This formulation collapses to two special cases. If all Φ's are equal to one, this form implies that instantaneous volatilities $\sigma_i(t)$ depend only on the time-to-expiry $T_{i-1} - t$, $\sigma_{i,\beta(t)} := \psi_{i-\beta(t)-1}$. We refer to this case as to the "time-to-expiry homogeneous piecewise constant" (TEHPC) formulation. If all ψ's collapse to one, instead, each forward rate has constant volatility $\sigma_i(t) = \Phi_i$, and we call this the "time-homogeneous piecewise constant" (THPC) formulation. These two extreme cases imply rather different evolutions in time for the term structure of volatilities, and the general SPC formulation usually ranges in-between, see again Chapter 6 of Brigo and Mercurio (2001) for the details.

Alternatively to PC volatilities, we may resort to a parametric linear-exponential (LE) formulation.

LE formulation

$$\boxed{\sigma_i(t) = \Phi_i \; \psi(T_{i-1} - t; a, b, c, d) =: \Phi_i \; \left([a(T_{i-1} - t) + d]e^{-b(T_{i-1}-t)} + c\right) \; .} \tag{3}$$

This parametric form leads to the same qualitative behaviour as the SPC one. In particular, it leads qualitatively to the same extreme cases as before when all Φ's are equal to one ("time-to-expiry homogeneous linear exponential" (TEHLE) formulation) or when a, b, d vanish with $c = 1$ (again the THPC formulation).

2.1 Caps and floors

Caps are collections of caplets. A T_{i-1}–*caplet* payoff pays at time T_i the amount

$$(F_i(T_{i-1}) - K)^+.$$

A T_{i-1}–caplet is thus a contract which pays at time T_i the difference between the T_i–maturity spot rate which has reset at time T_{i-i} and a strike rate K, if this difference is positive, and zero otherwise. Since in the LFM F_i follows a geometric Brownian motion under Q^i, the above expectation is computed immediately as a classical Black price. It follows that caps are priced easily in terms of instantaneous volatilities of the rates F. In particular, the volatility parameter $v_{T_{i-1}-\text{caplet}}\sqrt{T_{i-1}}$ entering Black's formula for the above caplet is the square root of the following average percentage variance

$$v^2_{T_{i-1}-\text{caplet}} = \frac{1}{T_{i-1}} \int_0^{T_{i-1}} \sigma_i(t)^2 dt, \tag{4}$$

multiplied by the expiry time T_{i-1}, which is straightforwardly computed on the basis of the parametric form chosen, be it PC or LE. For more details see Brigo and Mercurio (2001).

3 The swap market model (LSM)

Assume a unit notional. A (prototypical) interest rate swap (IRS) is a contract that exchanges payments between two differently indexed legs. At every instant T_j in a prespecified set of dates $T_{\alpha+1}, ..., T_\beta$ the fixed leg pays out an amount corresponding to a fixed interest rate K,

$$\tau_j K \ ,$$

whereas the floating leg pays an amount corresponding to the interest rate $F_i(T_{i-1})$ resetting at the previous instant T_{i-1} for the maturity given by the current payment instant T_i.

In this setting the floating leg rate resets at dates $T_\alpha, T_{\alpha+1}, \ldots, T_{\beta-1}$ and pays at dates $T_{\alpha+1}, \ldots, T_\beta$.

The *discounted* payoff of our IRS at a time $t < T_\alpha$ can be expressed either as

$$\sum_{i=\alpha+1}^{\beta} D(t, T_i)\ \tau_i(F_i(T_{i-1}) - K),$$

or as

$$D(t, T_\alpha) \sum_{i=\alpha+1}^{\beta} P(T_\alpha, T_i)\ \tau_i(F_i(T_\alpha) - K), \tag{5}$$

both expressions leading to the same risk-neutral expectation, giving the t-price of our IRS:

$$\sum_{i=\alpha+1}^{\beta} P(t, T_i)\tau_i(F_i(t) - K) \ = \ \sum_{i=\alpha+1}^{\beta} \left[P(t, T_{i-1}) - (1 + \tau_i K)P(t, T_i)\right].$$

From this formula we notice that, as is well known, neither volatility nor correlation of rates affect the value of an IRS.

The *forward swap rate* corresponding to the above IRS at time t is the particular value of the fixed-leg rate K that makes the contract fair, i.e. that makes its present value at time t equal to zero. We obtain, after straightforward manipulations,

$$S_{\alpha,\beta}(t) = \frac{1 - \prod_{j=\alpha+1}^{\beta} \frac{1}{1+\tau_j\ F_j(t)}}{\sum_{i=\alpha+1}^{\beta} \tau_i \prod_{j=\alpha+1}^{i} \frac{1}{1+\tau_j\ F_j(t)}}$$

so that the forward swap rate can be expressed in terms of spanning forward rates.

As an alternative expression we may write, through elementary manipulations

$$S_{\alpha,\beta}(t) = \sum_{i=\alpha+1}^{\beta} w_i(t)\ F_i(t), \tag{6}$$

$$w_i(t) = w_i(F_{\alpha+1}(t), F_{\alpha+2}(t), \dots, F_\beta(t)) = \frac{\tau_i \prod_{j=\alpha+1}^{i} \frac{1}{1+\tau_j\ F_j(t)}}{\sum_{k=\alpha+1}^{\beta} \tau_k \prod_{j=\alpha+1}^{k} \frac{1}{1+\tau_j\ F_j(t)}} .$$

This last expression is important because it can lead to useful approximations as follows. Formula (6) suggests that forward swap rates can be interpreted as weighted averages of spanning forward rates. However, notice carefully that the weights w's depend on the F's, so that we do not have a weighted average in strict sense. Based on empirical studies which show the variability of the w's to be small compared to the variability of the F's, one can approximate the w's by their (deterministic) initial values $w(0)$ and obtain

$$S_{\alpha,\beta}(t) \approx \sum_{i=\alpha+1}^{\beta} w_i(0)\ F_i(t)\ . \tag{7}$$

This can be helpful for example in estimating the absolute volatility of swap rates from the absolute volatility of forward rates.

Finally, notice that the IRS discounted payoff (5) for K different from the swap rate can be expressed also in terms of swap rates as

$$D(t, T_\alpha)\ (S_{\alpha,\beta}(T_\alpha) - K) \sum_{i=\alpha+1}^{\beta} \tau_i P(T_\alpha, T_i). \tag{8}$$

A swaption is an option to enter an IRS at a future time T_α. Clearly, this right will be exercised only when the swap rate at the exercise time T_α is larger than the IRS fixes rate (strike) K, so that the resulting IRS will have positive value. Consequently, the swaption payoff can be written as

$$D(t, T_\alpha)\ (S_{\alpha,\beta}(T_\alpha) - K)^+ \sum_{i=\alpha+1}^{\beta} \tau_i P(T_\alpha, T_i). \tag{9}$$

A numeraire under which the above swap rate follows a martingale is

$$C_{\alpha,\beta}(t) = \sum_{i=\alpha+1}^{\beta} \tau_i P(t, T_i).$$

Indeed, the product $C_{\alpha,\beta}(t)\ S_{\alpha,\beta}(t) = P(t,T_\alpha) - P(t,T_\beta)$ gives the price of a tradable asset which, expressed in C units, coincides with our forward swap rate. Therefore, when choosing C as numeraire, the forward swap rate $S_{\alpha,\beta}(t)$ evolves according to a martingale under the related measure $Q^{\alpha,\beta}$. Such measure is usually termed the "swap measure". The lognormal forward-swap model (LSM) assumes the following geometric-Brownian driftless dynamics for a forward-swap rate $S_{\alpha,\beta}(t)$ under its swap measure $Q^{\alpha,\beta}$:

$$d\ S_{\alpha,\beta}(t) = \sigma^{(\alpha,\beta)}(t) S_{\alpha,\beta}(t)\ dW_t^{\alpha,\beta},$$

where $W^{\alpha,\beta}$ is a standard Brownian motion under $Q^{\alpha,\beta}$. We denote by $v^2_{\alpha,\beta}(T)$ the average percentage variance of the forward swap rate in the interval $0, T$ times the interval length:

$$v^2_{\alpha,\beta}(T) = \int_0^T (\sigma^{(\alpha,\beta)}(t))^2\ dt\ .$$

When pricing a swaption (for example at time $t = 0$), this model is particularly convenient, since it yields a lognormal distribution for S and thus the well-known Black formula for swaptions, structurally analogous to the Black formula for caps:

$$\widetilde{E}\left(D(0,T_\alpha)\ (S_{\alpha,\beta}(T_\alpha) - K)^+\ C_{\alpha,\beta}(T_\alpha)\right) = C_{\alpha,\beta}(0) E^{\alpha,\beta}(S_{\alpha,\beta}(T_\alpha) - K)^+$$

$$= C_{\alpha,\beta}(0)\ \text{Black}(K, S_{\alpha,\beta}(0), v_{\alpha,\beta}(T_\alpha))$$

where "Black" denotes the core of Black's formula.

4 LFM vs LSM swap-rate distributions under the swap measure

We have seen in the previous section how one can price swaptions with the LSM, basically resorting to Black's formula. Now we consider instead pricing of swaptions with the LFM, and try to quantify the difference between the two models. The LSM swaption formula is based on a lognormal distribution for the swap rate $S_{\alpha,\beta}(T_\alpha)$ coming from the LSM dynamics under the swap measure $Q^{\alpha,\beta}$. In order for a comparison of distributions to make sense, we need both distributions under the same measure. Therefore we need the dynamics of forward Libor rates F under the LSM numeraire $C_{\alpha,\beta}(t)$. By applying the change of numeraire technique, we see that

$$\begin{aligned} dF_k(t) &= \sigma_k(t) F_k(t) \left(\mu_k^{\alpha,\beta}(t)dt + dZ_k(t)\right), \qquad (10) \\ \mu_k^{\alpha,\beta}(t) &= \sum_{j=\alpha+1}^{\beta} (2 \cdot 1_{(j \le k)} - 1)\tau_j \frac{P(t,T_j)}{C_{\alpha,\beta}(t)} \sum_{i=\min(k+1,j+1)}^{\max(k,j)} \frac{\tau_i \rho_{k,i} \sigma_i(t) F_i(t)}{1 + \tau_i F_i(t)}, \end{aligned}$$

where the Z's are now Brownian motions under $Q^{\alpha,\beta}$. This is a closed set of SDEs when k ranges from $\alpha+1$ to β, since the terms

$$\frac{P(t,T_j)}{C_{\alpha,\beta}(t)}$$

can be easily expressed as suitable functions of the spanning forward rates

$$F_{\alpha+1}(t),\ldots,F_\beta(t)\ .$$

By substituting the above forward rates into the previously derived algebraic expression defining swap rates in terms of forward rates,

$$S_{\alpha,\beta}(t) = \frac{1-\prod_{j=\alpha+1}^{\beta}\frac{1}{1+\tau_j\ F_j(t)}}{\sum_{i=\alpha+1}^{\beta}\tau_i\prod_{j=\alpha+1}^{i}\frac{1}{1+\tau_j\ F_j(t)}} \tag{11}$$

we obtain the **LFM swap rate under the swap measure**.

As an alternative to the LFM swap rate given by Equations (11,10) under the swap measure, we have seen earlier the **LSM swap rate under the swap measure** stemming from the lognormal dynamics

$$d\ S_{\alpha,\beta}(t) = \sigma^{(\alpha,\beta)}(t)S_{\alpha,\beta}(t)\ dW_t^{\alpha,\beta}. \tag{12}$$

Now, when computing the swaption price as the discounted expectation

$$\widetilde{E}\left(D(0,T_\alpha)\ (S_{\alpha,\beta}(T_\alpha)-K)^+\ C_{\alpha,\beta}(T_\alpha)\right) = C_{\alpha,\beta}(0)E^{\alpha,\beta}[(S_{\alpha,\beta}(T_\alpha)-K)^+]$$

we have two possibilities: We can do this either with the LFM given by Equations (11,10) or with the LSM given by Equation (12). The result is not the same.

Remark 4.1. *While the swap rate coming from the LSM dynamics (12) is lognormally distributed, the swap rate coming from the LFM dynamics (11,10) is not lognormal. This results in the two models being theoretically incompatible.*

Based on this remark, a quantity apt to measure the distance between the two models is the distance of the distribution of the LFM swap rate from the family of lognormal densities in which the distribution of the LSM swap rate lies.

In order to study this distance we need the distribution of the swap rate (11) under the swap measure obtained via simulation of (10). Therefore we now explain how the Monte Carlo simulation of equation (10) can be achieved.

Incidentally, while keeping in mind (11), notice carefully that the above expectation leading to the swaption price depends on the *joint* distribution of spanning forward rates, so that terminal correlations between different forward rates have an heavy impact on swaptions prices.

In order to simulate S through (11), we need to generate p realizations of these spanning forward rates according to the dynamics (10). The simulated F's allow us to evaluate the LFM swap rate S in each realization and back out its quantiles and therefore its distribution under the swap measure, although we will eventually use entropy estimators rather than empirical density reconstruction in order to estimate our distributional distance.

Since the dynamics (10) does not lead to a distributionally known process, we need to discretize such dynamics with a sufficiently (but not too) small time step Δt in order to reduce the random inputs to the distributionally known (Gaussian independent) shocks $Z(t+\Delta t) - Z(t)$.

In doing so taking logs can be helpful: By Itô's formula

$$d \ln F_k(t) = (\sigma_k(t)\ \mu_k^{\alpha,\beta}(t) - \tfrac{1}{2}\sigma_k(t)^2)\ dt + \sigma_k(t)\ dZ_k(t)\ . \tag{13}$$

This last equation has the advantage that the diffusion coefficient is deterministic. As a consequence, the naive Euler scheme coincides with the more sophisticated Milstein scheme, so that the discretization

$$\begin{aligned} \ln F_k^{\Delta t}(t+\Delta t) &= \ln F_k^{\Delta t}(t) + (\sigma_{k,\beta(t)}\ \mu_k^{\alpha,\beta}(t) - \tfrac{1}{2}\sigma_{k,\beta(t)}^2)\ \ \Delta t \\ &+ \int_t^{t+\Delta t} \sigma_k(t)\ dZ_k(t) \end{aligned} \tag{14}$$

leads to an approximation of the true process such that there exists a δ_0 with

$$E^{\alpha,\beta}\{|\ln F_k^{\Delta t}(T_\alpha) - \ln F_k(T_\alpha)|\} \le C(T_\alpha)(\Delta t)^1 \ \text{ for all } \ \Delta t \le \delta_0$$

where $C(T_\alpha)$ is a positive constant (strong convergence of order 1, Klöden and Platen (1995)). Recall that the (vector) integral term is normally distributed and easy to simulate, according to

$$\int_t^{t+\Delta t} \operatorname{diag}(\sigma(t))\ dZ(t) \sim \mathcal{N}(0, \Sigma(t)), \quad \Sigma(t)_{i,j} = \int_t^{t+\Delta t} \rho_{i,j}\sigma_i(u)\sigma_j(u)du.$$

Moreover, from the properties of Brownian motion, integrals on adjacent discretization intervals are independent, thus making the simulation immediate.

A possible problem concerns the instantaneous correlation matrix. Clearly, the full correlation matrix ρ features $M\ (M-1)/2$ parameters (where $M = \beta - \alpha$ is the number of forward rates), which can be a large number. Therefore, a parsimonious parametric form has to be found for ρ, based on a reduced number of parameters.

We start by reducing the number of noise factors. In general, one can choose an orthogonal (full rank) $M \times n$ matrix and replace dZ by BdW, where W is a n dimensional standard Brownian motion (in particular, $dWdW' = Idt$).

In doing so, we move from a noise correlation structure

$$dZdZ' = \rho\ dt$$

(where ρ in general is a $M \times M$ full rank matrix) to

$$BdW(BdW)' = BdWdW'B' = BB'dt.$$

Therefore, our new instantaneous noise-correlation matrix is $B \times B'$ whose rank is n.

We define

$$\rho^B = B \times B'.$$

A parametric form has to be chosen for B. Rebonato (1998) suggests the following general form for the i-th row of B:

$$\begin{aligned} b_{i,1} &= \cos\theta_{i,1} \\ b_{i,k} &= \cos\theta_{i,k}\sin\theta_{i,1}\cdots\sin\theta_{i,k-1}, \quad 1 < k < n, \\ b_{i,n} &= \sin\theta_{i,1}\cdots\sin\theta_{i,n-1}, \end{aligned} \tag{15}$$

for $i = 1, 2, \ldots, M$. Notice that with this parameterization clearly ρ^B is positive semidefinite and its diagonal terms are ones. It follows that ρ^B is a possible correlation matrix. The number of parameters in this case is $M \times (n-1)$, even though there is a canonical form under which $(M - n/2) \times (n-1)$ angles suffice, as pointed out for example in Rapisarda, Brigo and Mercurio (2002).

We will focus in this paper on a simple two-factor structure, $n = 2$, consisting of M parameters. This is obtained as

$$b_{i,1} = \cos\theta_{i,1} \quad b_{i,2} = \sin\theta_{i,1} \ . \tag{16}$$

We drop the second subscript for θ.

In such a case

$$\rho^B_{i,j} = b_{i,1}b_{j,1} + b_{i,2}b_{j,2} = cos(\theta_i - \theta_j), \tag{17}$$

which we will use in the following. This structure consists of M parameters $\theta_1, \ldots, \theta_M$.

Whichever correlation parameterization is chosen, the following step consists of calibrating the model to swaption prices, thus deducing the ρ^B from swaptions pricing. We will not deal with swaptions calibration in this paper, referring instead the interested reader to Brigo and Mercurio (2001).

Since here we are interested in expressing the distributional difference between the swap rates in the LSM and LFM, we need a method to measure the distance between distributions. This will be dealt with in two sections. Before facing this task, we report a formula connecting caplet volatilities to swaptions volatilities. This approximated formula is often used in practice to bridge the gap in-between the LFM and LSM without resorting to Monte Carlo simulations.

Recall the forward swap rate dynamics underlying the LSM, i.e. leading to Black's formula for swaptions:

$$d\ S_{\alpha,\beta}(t) = \sigma^{(\alpha,\beta)}(t) S_{\alpha,\beta}(t)\ dW_t^{\alpha,\beta}.$$

A crucial role in the LSM is played by the Black swap (squared) volatility (multiplied by time T_α)

$$v^2_{\alpha,\beta}(T_\alpha) = \int_0^{T_\alpha} \sigma^2_{\alpha,\beta}(t)dt = \int_0^{T_\alpha} (d\ln S_{\alpha,\beta}(t))(d\ln S_{\alpha,\beta}(t)) = \int_0^{T_\alpha} d\langle \ln S_{\alpha,\beta}, \ln S_{\alpha,\beta}\rangle_t$$

entering Black's formula for swaptions. It is possible to compute, under a number of approximations, an analogous quantity in the LFM.

Recall from Formula (7) that forward swap rates can be approximated as a weighted average of forward rates with constant frozen weights. By differentiating this equation, substituting the F's dynamics, taking quadratic variations, doing some more freezing and finally integrating, one finds the following approximation for $v^2_{\alpha,\beta}(T_\alpha)$ to be put in Black's formula for evaluating swaptions:

$$\begin{aligned}(v^{\text{LFM}}_{\alpha,\beta})^2 &= \int_0^{T_\alpha} (d\ln S_{\alpha,\beta}(t))(d\ln S_{\alpha,\beta}(t)) = \int_0^{T_\alpha} \langle \ln S_{\alpha,\beta}, \ln S_{\alpha,\beta}\rangle_t \\ &= \sum_{i,j=\alpha+1}^{\beta} \frac{w_i(0)w_j(0)F_i(0)F_j(0)\rho_{i,j}}{S_{\alpha,\beta}(0)^2} \int_0^{T_\alpha} \sigma_i(t)\sigma_j(t)\, dt\end{aligned} \tag{18}$$

(for a derivation and references see again Brigo and Mercurio (2001)).

Formula (18) is obtained under a number of approximations, and at first one would imagine its quality to be rather poor. However, it turns out that the approximation is not at all bad. A slightly more sophisticated version of this procedure has been pointed out for example by Hull and White (1999), but the differences between the two formulae are negligible in most situations. For a derivation and some numerical Monte Carlo tests see for example Chapter 8 of Brigo and Mercurio (2001).

5 Distance between distributions: The Kullback Leibler information

In this section we introduce briefly the Kullback-Leibler information and we explain its importance for our problem, see also Brigo and Hanzon (1998). Suppose we are given the space H of all the densities of probability measures on the real line equipped with its Borel field, which are absolutely continuous w.r.t. the Lebesgue measure. Then define

$$D(p_1, p_2) := E_{p_1}\{\log p_1 - \log p_2\} \geq 0, \quad p_1, p_2 \in H, \tag{19}$$

where in general

$$E_p\{\phi\} = \int \phi(x)p(x)dx, \quad p \in H.$$

The above quantity is the well-known Kullback-Leibler information (KLI). Its non-negativity follows from the Jensen inequality. It gives a measure of how much the density p_2 is displaced w.r.t. the density p_1. We remark the important fact that D is not a distance: in order to be a metric, it should be symmetric and satisfy the triangular inequality, which is not the case.

However, the KLI features many properties of a distance in a generalized geometric setting (see for instance Amari (1985)). For example, it is well-known that the KLI is infinitesimally equivalent to the Fisher information metric around every point of a finite–dimensional manifold of densities such as $EM(c)$ defined below. For this reason, we will refer to the KLI as to a "distance" even if it is not a metric.

Consider a finite dimensional manifold of exponential probability densities such as

$$\begin{aligned} EM(c) &= \{p(\cdot,\theta) : \theta \in \Theta \subset I\!R^m\}, \quad \Theta \text{ open in } I\!R^m, \\ p(\cdot,\theta) &= \exp[\theta_1 c_1(\cdot) + ... + \theta_m c_m(\cdot) - \psi(\theta)], \end{aligned} \tag{20}$$

expressed w.r.t the expectation parameters η defined by

$$\eta_i(\theta) = E_{p(\cdot,\theta)}\{c_i\} = \partial_{\theta_i}\psi(\theta), \quad i = 1,..,m \tag{21}$$

(see for example Amari (1985), Brigo (1999) or Brigo, Hanzon and Le Gland (1999) for more details on the geometry of exponential families).

We define $p(x;\eta(\theta)) := p(x,\theta)$ (the semicolon/colon notation identifies the parameterization).

Now suppose we are given a density $p \in H$, and we want to approximate it by a density of the finite dimensional manifold $EM(c)$. It seems then reasonable to find a density $p(\cdot,\theta)$ in $EM(c)$ which minimizes the Kullback Leibler information $D(p,.)$. Compute

$$\begin{aligned} \min_\theta D(p, p(\cdot,\theta)) &= \min_\theta \{E_p[\log p - \log p(\cdot,\theta)]\} \\ &= E_p \log p - \max_\theta \{\theta_1 E_p c_1 + ... + \theta_m E_p c_m - \psi(\theta)\} \\ &= E_p \log p - \max_\theta V(\theta), \\ V(\theta) &:= \theta_1 E_p c_1 + ... + \theta_m E_p c_m - \psi(\theta). \end{aligned}$$

It follows immediately that a necessary condition for the minimum to be attained at θ^* is

$\partial_{\theta_i} V(\theta^*) = 0, \quad i = 1,...,m$

which yields

$$E_p c_i - \partial_{\theta_i}\psi(\theta^*) = E_p c_i - E_{p(\cdot,\theta^*)} c_i = 0, \qquad i = 1,..,m$$

i.e. $E_p c_i = \eta_i(\theta^*), \ \ i = 1,..,m$. This last result indicates that according to the Kullback Leibler information, the best approximation of p in the manifold $EM(c)$ is given by the density of $EM(c)$ which shares the same c_i expectations (c_i-moments) as the given density p. This means that in order to approximate p we only need its c_i moments, $i = 1,2,..,m$.

The above discussion provides also a way to compute the distance of the density p from the exponential family $EM(c)$ as the distance between p and its projection $p(\cdot,\theta^*)$ onto $EM(c)$ in the KL sense. We have

$$\begin{aligned} D(p, EM(c)) &= E_p \log p \; - (\theta_1^* E_p c_1 + \ldots + \theta_m^* E_p c_m - \psi(\theta^*)) \\ &= E_p \log p \; - (\theta_1^* \eta_1(\theta^*) + \ldots + \theta_m^* \eta_m(\theta^*) - \psi(\theta^*)). \end{aligned} \tag{22}$$

One can look at the problem from the opposite point of view. Suppose we decide to approximate the density p by taking into account only its m c_i–moments. It can be proved (see Kagan, Linnik, and Rao (1973), Theorem 13.2.1) that the maximum entropy distribution which shares the c–moments with the given p belongs to the family $EM(c)$.

Summarizing: If we decide to approximate by using c–moments, then entropy analysis supplies arguments to use the family $EM(c)$; and if we decide to use the approximating family $EM(c)$, Kullback–Leibler says that the "closest" approximating density in $EM(c)$ shares the c–moments with the given density.

This moments-matching characterization of the projected density for exponential families is the main reason why we resort to the KLI as a "distance" between distribution. Alternatively, we might use the Hellinger distance, which is defined, for two densities $p_1, p_2 \in H$ as

$$H(p_1, p_2) := 2 - 2 \int \sqrt{p_1(x) p_2(x)} dx, \tag{23}$$

from which we see that the HD takes values in $[0, 2]$ and is a real metric. It is well-known, however, that the KLI is infinitesimally equivalent to the Hellinger distance around every point of a finite–dimensional manifold of densities such as $EM(c)$ defined above. For this reason one refers to the KLI as to a "distance" even if it is not a metric. Indeed, consider the two densities $p(\cdot, \theta)$ and $p(\cdot, \theta + d\theta)$ of $EM(c)$. By expanding in Taylor series, we obtain easily

$$\begin{aligned} K(p(\cdot, \theta), p(\cdot, \theta + d\theta)) &= -\sum_{i=1}^{m} E_{p(\cdot,\theta)} \{ \frac{\partial \log p(\cdot, \theta)}{\partial \theta_i} \} \, d\theta_i \\ &\quad - \sum_{i,j=1}^{m} E_{p(\cdot,\theta)} \{ \frac{\partial^2 \log p(\cdot, \theta)}{\partial \theta_i \partial \theta_j} \} \, d\theta_i \, d\theta_j + O(|d\theta|^3) \end{aligned}$$

which is the same expression we obtain by expanding $H(p(\cdot, \theta), p(\cdot, \theta + d\theta))$. Given this first-order relationship, we expect that the Hellinger distance would lead us to the same results as the KLI, since the KLI distances we will find are rather small.

6 Distance of the LFM swap rate from the lognormal family of distributions

Since under the swap measure the LSM dynamics for the swap rate follows a (driftless) geometric Brownian motion, we consider here the SDE describing a general geometric Brownian motion

$$dS_t = \; \mu(t) S_t \; dt + \; \sigma(t) \; S_t \; dW_t \; , \quad S_0 = s_0$$

whose solution is

$$S_t = s_0 \exp\left[\int_0^t (\mu(u) - \tfrac{1}{2}\sigma^2(u))du + \int_0^t \sigma(u)dW_u\right],$$

so that

$$\log S_t \sim \mathcal{N}\left(\log s_0 + \int_0^t (\mu(u) - \tfrac{1}{2}\sigma^2(u))du, \int_0^t \sigma^2(u)du\right) . \tag{24}$$

The probability density p_{S_t} of S_t, at any time t, is therefore given by

$$\begin{aligned} p_{S_t}(x) &= p(x,\theta(t)) = \exp\left\{\theta_1(t)\ln\frac{x}{s_0} + \theta_2(t)\ln^2\frac{x}{s_0} - \psi(\theta_1(t),\theta_2(t))\right\}, \\ \theta_1(t) &= \frac{\int_0^t \mu(u)\,du}{\int_0^t \sigma^2(u)du} - \frac{3}{2}, \quad \theta_2(t) = -\frac{1}{2\int_0^t \sigma^2(u)du}, \quad \psi(\theta_1(t),\theta_2(t)) = -\frac{(\theta_1(t)+1)^2}{4\theta_2(t)} + \tfrac{1}{2}\ln\left(\frac{-\pi\ s_0^2}{\theta_2(t)}\right), \end{aligned}$$

where $x > 0$, and is clearly in the exponential class, with $c_1(x) = \ln(x/s_0)$, $c_2 = c_1^2$. We will denote by $\mathcal{L}$ the related exponential family $EM(c)$. As concerns the expectation parameters for this family, they are readily computed as follows:

$$\begin{aligned} \eta_1 &= E_\theta \ln(x/s_0) = \partial_{\theta_1}\psi(\theta_1,\theta_2) = -\frac{\theta_1+1}{2\theta_2} \\ \eta_2 &= E_\theta \ln^2(x/s_0) = \partial_{\theta_2}\psi(\theta_1,\theta_2) = \left(\frac{\theta_1+1}{2\theta_2}\right)^2 - \frac{1}{2\theta_2} . \end{aligned}$$

As for the Gaussian family, in this particular family the θ parameters can be computed back from the η parameters by inverting the above formulae:

$$\begin{aligned} \theta_1 &= \frac{\eta_1}{\eta_2 - \eta_1^2} - 1 , \qquad (25) \\ \theta_2 &= -\frac{1}{2(\eta_2 - \eta_1^2)} , \\ \psi(\theta_1,\theta_2) &= \tfrac{1}{2}\left[\frac{\eta_1^2}{\eta_2 - \eta_1^2} + \ln(2\pi(\eta_2 - \eta_1^2)s_0^2)\right] . \end{aligned}$$

We can now compute the distance of a density p from the lognormal family $\mathcal{L}$ by applying formula (22):

$$D(p,\mathcal{L}) = E_p \ln p\ - (\theta_1^*\eta_1(\theta^*) + \theta_2^*\eta_2(\theta^*) - \psi(\theta^*)),$$

where, as previously seen, minimizing the distance implies finding the parameters θ^* such that

$$\eta_1(\theta^*) = E_p \ln(x/s_0), \quad \eta_2(\theta^*) = E_p \ln^2(x/s_0) .$$

By substituting (25), omitting the argument θ^* and simplifying, we obtain

$$D(p,\mathcal{L}) = E_p \ln p\ + \tfrac{1}{2} + \eta_1 + \tfrac{1}{2}\ln(2\pi(\eta_2 - \eta_1^2)s_0^2) .$$

Actually, the s_0 term is kind of redundant when computing the distance. We can thus resort to the simpler moments

$$\bar{\eta}_1(\bar{\theta}^*) = E_p \ln(x), \quad \bar{\eta}_2(\bar{\theta}^*) = E_p \ln^2(x) \ ,$$

and compute the distance as

$$D(p, \mathcal{L}) = E_p \ln p \ + \tfrac{1}{2} + \bar{\eta}_1 + \tfrac{1}{2} \ln(2\pi(\bar{\eta}_2 - \bar{\eta}_1{}^2)) \ . \tag{26}$$

As noticed before, the LSM swap–rate density under the swap measure belongs to the family $\mathcal{L}$: Such density is given by $p(\cdot, \theta(T_\alpha))$ above when taking $s_0 = S_{\alpha,\beta}(0)$, $\mu(t) = 0$ and $\sigma(t) = \sigma_{\alpha,\beta}(t)$.

Now consider instead the LFM swap rate under the swap measure, obtained once again through (11) and (10). This second swap rate will not be lognormally distributed. Let $p_{\alpha,\beta}$ denote the probability density of the LFM swap rate $S_{\alpha,\beta}(T_\alpha)$ under the swap measure $Q^{\alpha,\beta}$.

We plan to compute numerically the distance of the LFM swap density $p_{\alpha,\beta}$ from the lognormal exponential family $\mathcal{L}$ where the LSM swap density lies. But we also plan to verify the volatility–approximation provided by the quantity $v^{\text{LFM}}_{\alpha,\beta}$ as follows: The density $p(\cdot, \theta^*) = p(\cdot; \eta(\theta^*))$ represents the lognormal density which is closest (in the Kullback-Leibler sense) to the swap-rate density $p_{\alpha,\beta}$ implied by the LFM under the swap measure. Incidentally, we can compute the terminal volatility implied by this lognormal density as

$$v^{\text{KLI}}_{\alpha,\beta} = \sqrt{\eta_2(\theta^*) - \eta_1(\theta^*)^2} = \sqrt{\bar{\eta}_2 - \bar{\eta}_1^2} \ .$$

This is the best approximation of the volatility of the swap rate based on the lognormal approximation. It can be interesting to compare such best approximation to the much handier approximation $v^{\text{LFM}}_{\alpha,\beta}$ considered earlier in (18).

Recall that $v^{\text{LFM}}_{\alpha,\beta}$ is obtained by "ignore the drifts" and "freeze stochastic coefficients" arguments, whereas $v^{\text{KLI}}_{\alpha,\beta}$ is obtained by minimizing the distance from the lognormal densities. Should the two results be close, this would represent a further confirmation of the validity of the industry formula $v^{\text{LFM}}_{\alpha,\beta}$.

Now we proceed by applying formula (26) according to the following scheme:

1. Simulate p realizations of the forward rates

$$F_{\alpha+1}(T_\alpha), F_{\alpha+2}(T_\alpha), \ldots, F_\beta(T_\alpha)$$

   under the swap measure $Q^{\alpha,\beta}$ through the discretized dynamics (14) with a sufficiently small time step;

2. Compute p realizations of the swap rate $S_{\alpha,\beta}(T_\alpha)$ of the LFM under the swap measure $Q^{\alpha,\beta}$ through (11) applied to each realization of the forward rates F vector obtained in the previous point;

3. Based on the simulated $S_{\alpha,\beta}(T_\alpha)$, compute firstly

$$\bar{\eta}_1 = E^{\alpha,\beta} \ln(S_{\alpha,\beta}(T_\alpha)), \quad \bar{\eta}_2 = E^{\alpha,\beta} \ln^2(S_{\alpha,\beta}(T_\alpha)) \ , v^{\text{KLI}}_{\alpha,\beta} = \sqrt{\bar{\eta}_2 - \bar{\eta}_1{}^2} \ ,$$

and secondly an approximation of

$$E_p \ln p := \int (\ln p_{\alpha,\beta}(x)) p_{\alpha,\beta}(x) dx \ .$$

This quantity is the opposite of entropy and can be estimated from the simulated $S_{\alpha,\beta}(T_\alpha)$'s through an entropy estimator. For example, Vasicek's (1976) estimator reads, in our case,

$$H_V(q,p) = -\frac{1}{p}\sum_{i=1}^{p} \ln \left[\frac{p}{2q} (S_{\alpha,\beta}(T_\alpha)_{[i+q]} - S_{\alpha,\beta}(T_\alpha)_{[i-q]}) \right],$$

where $S_{\alpha,\beta}(T_\alpha)_{[j]}$ is the j-th order statistics from our sample. We set $S_{\alpha,\beta}(T_\alpha)_{[j]} = S_{\alpha,\beta}(T_\alpha)_{[1]}$ for $j < 1$ and $S_{\alpha,\beta}(T_\alpha)_{[j]} = S_{\alpha,\beta}(T_\alpha)_{[p]}$ for $j > p$.

This estimator converges to the desired integral in probability as

$$p \to \infty, \quad q \to \infty, \quad q/p \to 0.$$

However, in our numerical simulations we considered a different type of entropy estimator. We used the plug-in estimate of entropy based on a cross-validation density estimate, proposed by Ivanov and Rozhkova (1981). For an overview of entropy estimators see for example Dudewicz and van der Meulen (1987). The estimator we used can be briefly summarized as follows. Let $S^1_{\alpha,\beta}(T_\alpha), \dots, S^p_{\alpha,\beta}(T_\alpha)$ be i.i.d. sample of swap rates with unknown probability density function $p_{\alpha,\beta}(x)$, and consider

$$p^{p,i}_{\alpha,\beta}(S^i_{\alpha,\beta}(T_\alpha)) = \frac{1}{pa_p}\sum_{j\neq i} K\left(\frac{S^i_{\alpha,\beta}(T_\alpha) - S^j_{\alpha,\beta}(T_\alpha)}{a_p}\right),$$

where $\{a_p\}$ satisfies the condition that $a_p \to 0$, $pa_p \to \infty$, and K is a kernel function. Note that we used Gaussian kernel in our computations. The estimator of Ivanov and Rozhkova (1981) can be written in the following form:

$$H_{IR}(p) = -\frac{1}{p}\sum_{i=1}^{p} \left\{ \ln p^{p,i}_{\alpha,\beta}(S^i_{\alpha,\beta}(T_\alpha)) \right\} I_{\left[S^i_{\alpha,\beta}(T_\alpha) \in A_p\right]}, \tag{27}$$

where with the set A_p one typically excludes the small and the tail values of $p^{p,i}_{\alpha,\beta}(S^i_{\alpha,\beta}(T_\alpha))$. Ivanov and Rozhkova (1981) showed that under certain conditions on K, $p_{\alpha,\beta}$, a_p and A_p, $H_{IR}(p)$ converges with probability 1 to the desired integral as $p \to \infty$. We also tried some alternative estimators, such as (4) and (6) in Miller (2003). This did not change our numerical results significantly.

4. With the quantities obtained from the previous point apply formula (26) and obtain $D(p_{\alpha,\beta}, \mathcal{L})$.

5. Compute $v^{\text{LFM}}_{\alpha,\beta}$ through formula (18) and compare this to $v^{\text{KLI}}_{\alpha,\beta}$ obtained above.

It is interesting to plot $D(p_{\alpha,\beta}, \mathcal{L})$ and the difference $|v^{\text{KLI}}_{\alpha,\beta} - v^{\text{LFM}}_{\alpha,\beta}|$, as α and β change, and also analyze these quantities in the different formulations of instantaneous volatilities and correlations. For each formulation, which are the parameters to which the distance is more sensitive?

7 Monte Carlo tests for measuring KLI

In this section we numerically test how the KLI "distance" between a simulated swap rate density and the family of lognormal densities changes for different parameterizations. We will consider a family of forward rates whose expiry/maturity pairs are $T_0 = 1y$, $T_1 = 2y$ up to $T_{18} = 19y$, and our initial input is

$$F_{0\div 9}(0) = [4.69 \;\; 5.01 \;\; 5.60 \;\; 5.84 \;\; 6.00 \;\; 6.13 \;\; 6.28 \;\; 6.27 \;\; 6.29 \;\; 6.23]/100$$

$$F_{10\div 19}(0) = [6.30 \;\; 6.36 \;\; 6.43 \;\; 6.48 \;\; 6.53 \;\; 6.40 \;\; 6.30 \;\; 6.18 \;\; 6.07 \;\; 5.94]/100,$$

$F_0(0)$ being the initial one-year spot rate. These values are consistent with the volatility and correlation values below, in that all such initial inputs will reflect a possible calibration of the LFM to caps and swaptions. The corresponding swap rates we consider are $S_{5,10}(0) = 0.06238$, $S_{10,15}(0) = 0.06411$ $S_{15,20}(0) = 0.06191$, $S_{5,15}(0) = 0.06312$, $S_{10,20}(0) = 0.06318$, $S_{5,20}(0) = 0.06283$. We adopt the LE formulation (3) for instantaneous volatilities; in some cases we will let it collapse to the TEHLE and THPC formulations respectively. For instantaneous correlations we resort to the angle form (17) in the parameters $\theta = [\,\theta_1 \,\ldots\, \theta_{19}]$. The values of the parameters a, b, c, d and the values of the θ's in the general case of the full LE formulation have been built so as to reflect possible joint calibrations of the LFM to caplets and swaptions. Such values are reported in points (2.a–c) below. For a discussion on which forms are to be preferred from the point of view of realistic behaviour of future volatilty structures (typically the forms of cases (2.a–c) below) see chapter 6 of Brigo and Mercurio (2001). A short map of our sets of testing parameters is given in the following.

(1.a) *Constant instantaneous (THPC) volatilities, typical rank-two correlations.*

LE formulation with $a = 0,\ b = 0,\ c = 1,\ d = 0$. This is actually the THPC formulation, since the ψ-part of the LE formulation is collapsed to one. We set

$$\begin{aligned} \Phi_{1\div 9} &= [\,0.1490\;0.1589\;0.1533\;0.1445\;0.1356\;0.1267\;0.1215\;0.1176\;0.1138\,], \\ \Phi_{10\div 19} &= [\,0.1106\;0.1076\;0.1046\;0.1017\;0.0989\;0.0978\;0.0974\;0.0969\;0.0965\;0.0961\,]. \end{aligned}$$

The correlation angles are taken as

$$\begin{aligned}\theta_{1\div 9} &= [\ 0.0147\ 0.0643\ 0.1032\ 0.1502\ 0.1969\ 0.2239\ 0.2771\ 0.2950\ 0.3630\ ],\\ \theta_{10\div 19} &= [\ 0.3810\ 0.4217\ 0.4836\ 0.5204\ 0.5418\ 0.5791\ 0.6496\ 0.6679\ 0.7126\ 0.7659\ ].\end{aligned}$$

This set of angles implies positive and decreasing instantaneous correlations when moving away from the "1" diagonal entries along columns.

(1.b) *Constant instantaneous (THPC) volatilities, perfect correlation.*

LE formulation with a, b, c, d and Φ's as in (1.a) and $\theta = [\ 0\ 0\ \ldots\ 0\ 0\ ]$, implying that all instantaneous correlations are set to one.

(1.c) *Constant inst. (THPC) volatilities, some negative rank-two correlations*

LE formulation with a, b, c, d and Φ's as in (1.a) and $\theta = [\ \theta_{1\div 9},\ \theta_{10\div 19}\ ]$, where

$$\begin{aligned}\theta_{1\div 9} &= [\ 0\ 0.0000\ 0.0013\ 0.0044\ 0.0096\ 0.0178\ 0.0299\ 0.0474\ 0.0728\ ],\\ \theta_{10\div 19} &= [\ 0.1100\ 0.1659\ 0.2534\ 0.3989\ 0.6565\ 1.1025\ 1.6605\ 2.0703\ 2.2825\ 2.2260\ ].\end{aligned}$$

These parameters θ imply some negative correlations while maintaining a decreasing correlation pattern when moving away from the diagonal in the resulting correlation matrix.

(2.a) *Humped and expiry-adjusted (LE) instantaneous volatilities depending only on time to expiry, typical rank-two correlations.*

LE formulation with $a = 0.1908,\ b = 0.9746,\ c = 0.0808,\ d = 0.0134$ and θ as in (1). This is the "most normal" situation, in that it reflects a joint calibration to caplets and swaptions volatilities. The parameters Φ's are, in our case:

$$\begin{aligned}\Phi_{1\div 9} &= [\,1.0500\ 1.0900\ 1.1025\ 1.1025\ 1.0913\ 1.0669\ 1.0624\ 1.0611\ 1.0544\,],\\ \Phi_{10\div 19} &= [\,1.0475\ 1.0386\ 1.0270\ 1.0132\ 0.9975\ 0.9979\ 1.0033\ 1.0079\ 1.0119\ 1.0152\,].\end{aligned}$$

(2.b) *Humped and maturity-adjusted (LE) instantaneous volatilities depending only on time to expiry, perfect correlation.*

LE formulation with a, b, c, d and Φ's as in (2.a) and $\theta = [\ 0\ 0\ \ldots\ 0\ 0\ ]$.

(2.c) *Humped and maturity-adjusted (LE) instantaneous volatilities depending only on time to expiry, some negative rank-two correlations.*

LE formulation with a, b, c, d and Φ's as in (2.a) and θ as in (1.c).

We now present the results obtained in evaluating the KLI for swap rates through the Monte Carlo method with antithetic shocks and with $2{\times}100000$ paths.

In the tables, we denote by *KLI* the estimated value of (26) and by *absdiff* the absolute differences $|v^{\text{KLI}}_{\alpha,\beta} - v^{\text{LFM}}_{\alpha,\beta}|$. First we would like to have a feeling for what it means to have a KLI distance of 0.006 between two distributions. We may resort to the KLI distance of two lognormals, which is easily computed analytically. Indeed, if we call θ_1, θ_2 the parameters of the first lognormal density (with corresponding expectation parameters η_1 and η_2) and $\hat{\theta}_1, \hat{\theta}_2$ the parameters of the second lognormal density (with corresponding expectation parameters $\hat{\eta}_1$ and $\hat{\eta}_2$), it is easy to compute the KLI distance as $(\hat{\theta}_1 - \theta_1)\hat{\eta}_1 + (\hat{\theta}_2 - \theta_2)\hat{\eta}_2 + \psi(\theta) - \psi(\hat{\theta})$. If we take for example $\eta_1 = \hat{\eta}_1 = 0.06$ (same mean) and then $\eta_2 = 0.04, \hat{\eta}_2 = 0.0404$, corresponding to $\theta_1 = 0.6484$, $\theta_2 = -27.4725$, $\hat{\theta}_1 = 0.6304$, $\hat{\theta}_2 = -27.1739$, we find a KLI distance of 0.00606, comparable in size to our distances above. Compute the standard deviations (volatilities) of the two distributions according to $\sqrt{\eta_2 - \eta_1^2}$. One has $\sqrt{\eta_2 - \eta_1^2} = 0.1908$, $\sqrt{\hat{\eta}_2 - \hat{\eta}_1^2} = 0.1918$. Recall that the two lognormal densities have the same mean of 0.06. Therefore, in a lognormal world with the mean fixed at 0.06, a KLI distance of 0.006 would amount to an absolute difference in volatility of about 0.001 for volatilities ranging around 0.19. This amounts to a percentage difference in standard deviations of 0.55%. This gives a feeling for the size of the distributional discrepancy our distances imply in the *worst* case we obtain from our simulations.

Before presenting our final results, a remark is in order on the accuracy of the KLI "distances" obtained by simulation. Typically, the standard error ranges about $3E-5$ with 200000 paths for the first distance 0.0001857 in Table (1), and for all other cases varies roughly in the same proportion with respect to the distance, so that for example the distance of 0.0068614 of Table (3) has a standard error of about 1E-3, and so on. Every time the standard error makes the distance pattern uncertain we add a question mark in the summary tables.

Fortunately, this happens only in five of the eighteen cases we analyzed, namely (i) in case (1.a) with $\uparrow \beta$, (ii) in case (1.c) with $\uparrow (\alpha, \beta)$, (iii) in case (2.a) with $\uparrow (\alpha, \beta)$, (iv) in case (2.a) with $\uparrow \beta$ and (v) in case (2.c) with $\uparrow (\alpha, \beta)$.

Now let us consider our test results, starting from Table 1 for case (1.a).

Swap	*KLI*	$v^{\text{KLI}}_{\alpha,\beta}/\sqrt{T_\alpha}$	$v^{\text{LFM}}_{\alpha,\beta}/\sqrt{T_\alpha}$	*absdiff*
$S_{5,10}(5)$	0.0001857	0.12376	0.12360	0.00016
$S_{10,15}(10)$	0.0000357	0.10516	0.10512	0.00004
$S_{15,20}(15)$	0.0000638	0.09705	0.09682	0.00023
$S_{5,10}(5)$	0.0001857	0.12376	0.12360	0.00016
$S_{5,15}(5)$	0.0002088	0.11466	0.11509	0.00043
$S_{5,20}(5)$	0.0003604	0.10951	0.10985	0.00034
$S_{5,20}(5)$	0.0003604	0.10951	0.10985	0.00034
$S_{10,20}(10)$	0.0001105	0.10150	0.10116	0.00034
$S_{15,20}(15)$	0.0000638	0.09705	0.09682	0.00023

Table 1: Results for case (1.a)

First consider distances from the lognormal family for $S_{5,10}(5), S_{10,15}(10)$, and $S_{15,20}(15)$ respectively (increasing maturity T_α, constant tenor $T_\beta - T_\alpha$, denoted "$\uparrow (\alpha, \beta)$"). The distance first decreases and then increases, displaying a "V" shape when plotted for example against α. This is interesting and might be due to the shape of the instantaneous volatility functions $\sigma(t)$ in this formulation. If not for particular shapes of volatilities, one would expect instead the distance to increase with the maturity, since the more the "non-lognormal dynamics" goes on, the more it is likely that one moves away from the lognormal distribution. Then consider $S_{5,10}(5), S_{5,15}(5)$, and $S_{5,20}(5)$ (increasing tenor, constant maturity, denoted "$\uparrow \beta$"). The distance increases this time, as expected, although the first two distances differ less than the standard error, so that we may not exclude a "V" shape a priori. Anyway, an increasing pattern is to be expected: if the tenor increases we are adding more forward rates to form the swap rate, and intuitively we move farther away from the lognormal family. Consider also $S_{5,20}(5), S_{10,20}(10)$, and $S_{15,20}(15)$ (increasing maturity, decreasing tenor, denoted "$\uparrow \alpha$"). The distance decreases, as is partly expected by the fact that at the final time we are adding less forward rates, even though the dynamics is propagated for longer times. Finally consider $S_{10,15}(10)$ and $S_{10,20}(10)$ (again increasing tenor, constant maturity). This time the distance increases, as expected.

At this point, as from (1.b), we set all correlations to one and recompute all distances. Table 2 shows our results in this case. The only pattern that has changed qualitatively concerns increasing maturity and decreasing tenor, denoted "$\uparrow \alpha$", which is now V-shaped instead of decreasing as before. Enlarging instantaneous correlation between far rates has caused the distance to increase with maturity in the final step.

Swap	*KLI*	$v^{\text{KLI}}_{\alpha,\beta}/\sqrt{T_\alpha}$	$v^{\text{LFM}}_{\alpha,\beta}/\sqrt{T_\alpha}$	*absdiff*
$S_{5,10}(5)$	0.0002073	0.12405	0.12380	0.00025
$S_{10,15}(10)$	0.0000178	0.10504	0.10530	0.00026
$S_{15,20}(15)$	0.0001339	0.09741	0.09701	0.00040
$S_{5,10}(5)$	0.0002073	0.12405	0.12380	0.00025
$S_{5,15}(5)$	0.0002814	0.11552	0.11583	0.00031
$S_{5,20}(5)$	0.0003917	0.11124	0.11143	0.00019
$S_{5,20}(5)$	0.0003917	0.11124	0.11143	0.00019
$S_{10,20}(10)$	0.0000370	0.10220	0.10186	0.00034
$S_{15,20}(15)$	0.0001339	0.09741	0.09701	0.00040

Table 2: Results for case (1.b)

Lowering now correlations again, including some negative entries in the correlation matrix, according to (1.c), gives us the results of Table 3. We see that negative correlations give the same qualitative results as positive correlations in (1.a), with the exception of the first pattern "$\uparrow (\alpha, \beta)$" which is now decreasing or humped (uncertainty coming from the standard error size). At this point, with negative correlation, this pattern is more counter-intuitive. The only reason for a decreasing distance of the swap rate distribution from the lognormal family when propagating a "non lognormal dynamics" for longer times and with the same tenor is given by lower $\sigma(\cdot)$ functions. Indeed, with the volatility formulation of cases (1.a)–(1.c) the term structure of future volatilities decreases considerably in time, so that farther forward rates have much lower volatilities than forward rates involved in "earlier" swap rates. For example, the swap rate $S_{15,20}$ involves F_{19}, whose volatility in (1.a)–(1.c) is set to 0.0961, whereas earlier swap rates may involve F_2, whose volatility is much higher and set to 0.1589. So even if the dynamics propagates for longer times, it does so with lower randomness, and this effect dominates the other one.

Swap	*KLI*	$v^{\text{KLI}}_{\alpha,\beta}/\sqrt{T_\alpha}$	$v^{\text{LFM}}_{\alpha,\beta}/\sqrt{T_\alpha}$	*absdiff*
$S_{5,10}(5)$	0.0001589	0.12402	0.12377	0.00025
$S_{10,15}(10)$	0.0001126	0.10348	0.10343	0.00005
$S_{15,20}(15)$	0.0000609	0.08767	0.08735	0.00032
$S_{5,10}(5)$	0.0001589	0.12402	0.12377	0.00025
$S_{5,15}(5)$	0.0002497	0.11370	0.11407	0.00037
$S_{5,20}(5)$	0.0068614	0.08629	0.08720	0.00091
$S_{5,20}(5)$	0.0068614	0.08629	0.08720	0.00091
$S_{10,20}(10)$	0.0023644	0.07192	0.07161	0.00031
$S_{15,20}(15)$	0.0000609	0.08767	0.08735	0.00032

Table 3: Results for case (1.c)

As a summary of patterns for the cases with constant instantaneous volatilities we display in Table 4 the behavior of the distance for different correlation configurations (1.a)–(1.c).

Action	Positive correlations	Perfect correlations	Some negative correl.
$\uparrow (\alpha, \beta)$	V shaped	V shaped	decreasing (humped?)
$\uparrow \beta$	increasing (V-shaped?)	increasing	increasing
$\uparrow \alpha$	decreasing	V shaped	decreasing

Table 4: Distance patterns against (α, β), β and α respectively, for cases (1.a), (1.b), (1.c)

Now let us move to commenting our results for cases (2.a)–(2.c), given in Tables 5, 6, and 7.

These results are qualitatively analogous to the results of the corresponding cases (1.a)–(1.c), with one strong exception and two weaker exceptions. The strong exception concerns the pattern "$\uparrow (\alpha, \beta)$" for the typical rank two correlations (case (2.a)), where we have an opposite humped pattern with respect to the earlier V-shaped case. This is due to the different volatility structure, that is now homogeneous with respect to time-to-maturity. Consider, however, that

Swap	*KLI*	$v^{\text{KLI}}_{\alpha,\beta}/\sqrt{T_\alpha}$	$v^{\text{LFM}}_{\alpha,\beta}/\sqrt{T_\alpha}$	*absdiff*
$S_{5,10}(5)$	0.00016634	0.11033	0.11017	0.00016
$S_{10,15}(10)$	0.00017273	0.09541	0.09534	0.00007
$S_{15,20}(15)$	0.00009037	0.08976	0.08969	0.00007
$S_{5,10}(5)$	0.00016634	0.11033	0.11017	0.00016
$S_{5,15}(5)$	0.00041173	0.09778	0.09803	0.00025
$S_{5,20}(5)$	0.00047629	0.09306	0.09320	0.00014
$S_{5,20}(5)$	0.00047629	0.09306	0.09320	0.00014
$S_{10,20}(10)$	0.00023379	0.08921	0.08895	0.00026
$S_{15,20}(15)$	0.00009037	0.08976	0.08969	0.00007

Table 5: Results for case (2.a)

Swap	*KLI*	$v^{\text{KLI}}_{\alpha,\beta}/\sqrt{T_\alpha}$	$v^{\text{LFM}}_{\alpha,\beta}/\sqrt{T_\alpha}$	*absdiff*
$S_{5,10}(5)$	0.00021961	0.11052	0.11035	0.00017
$S_{10,15}(10)$	0.00007206	0.09554	0.09552	0.00002
$S_{15,20}(15)$	0.00015586	0.09003	0.08987	0.00016
$S_{5,10}(5)$	0.00021961	0.11052	0.11035	0.00017
$S_{5,15}(5)$	0.00044884	0.09848	0.09867	0.00019
$S_{5,20}(5)$	0.00057553	0.09453	0.09457	0.00004
$S_{5,20}(5)$	0.00057553	0.09453	0.09457	0.00004
$S_{10,20}(10)$	0.00011407	0.08982	0.08957	0.00025
$S_{15,20}(15)$	0.00015586	0.09003	0.08987	0.00016

Table 6: Results for case (2.b)

in this case we have uncertainty in the pattern due to the standard error, and that in fact the pattern in (2.a) could be decreasing.

The first weaker exception concerns the pattern "$\uparrow (\alpha, \beta)$" for the case with some negative correlations (case (2.c)), where we have a humped pattern instead of a decreasing one, although both patterns are uncertain, with the possibility of the patterns coinciding in a decreasing or humped configuration. The second possible weaker exception is for $\uparrow \beta$ with positive correlation (case (2.a)), where the two patterns coincide unless the standard error changes them in two opposite configurations.

As a summary we display in Table 8 the behavior of the distance for different correlation configurations in cases (2).

As an example for visualizing the distance of a swap rate density from the best approximating lognormal density in the KLI sense, we have chosen the two swap rates that lead to the largest and to the smallest distance over all cases considered. We show in Figure 1 the Monte

Swap	*KLI*	$v^{\text{KLI}}_{\alpha,\beta}/\sqrt{T_\alpha}$	$v^{\text{LFM}}_{\alpha,\beta}/\sqrt{T_\alpha}$	*absdiff*
$S_{5,10}(5)$	0.00017029	0.11050	0.11032	0.00018
$S_{10,15}(10)$	0.00017182	0.09392	0.09388	0.00004
$S_{15,20}(15)$	0.00008607	0.08101	0.08078	0.00023
$S_{5,10}(5)$	0.00017029	0.11050	0.11032	0.00018
$S_{5,15}(5)$	0.00046520	0.09670	0.09721	0.00051
$S_{5,20}(5)$	0.00635040	0.07363	0.07409	0.00046
$S_{5,20}(5)$	0.00635040	0.07363	0.07409	0.00046
$S_{10,20}(10)$	0.00265217	0.06370	0.06344	0.00026
$S_{15,20}(15)$	0.00008607	0.08101	0.08078	0.00023

Table 7: Results for case (2.c)

Action	Positive correlation	Perfect correlation	Some negative correl.
$\uparrow (\alpha, \beta)$	humped (decreasing?)	V shaped	humped (decreasing?)
$\uparrow \beta$	increasing (humped?)	increasing	increasing
$\uparrow \alpha$	decreasing	V shaped	decreasing

Table 8: Distance patterns against (α, β), β and α respectively, for cases (2.a), (2.b), (2.c)

Carlo simulated densities for the swap rates $S_{5,20}(5)$ and $S_{10,15}(10)$ versus their lognormal KLI projections in case the parameters are as in points (1.c) and (1.b) respectively.

We notice also that *absdiff* is always small, meaning that the industry approximation $v^{\text{LFM}}_{\alpha,\beta}$ is good since it is always close to $v^{\text{KLI}}_{\alpha,\beta}$, i.e. to the best one can do with a lognormal family.

8 Conclusions

Our KLI analysis confirms that swap rates associated with the LIBOR market model are close to being log normal. This has been checked via a distributional distance obtained through Monte Carlo simulation. Our analysis also confirms the goodness of the standard market approximation for swaption volatilities in the LIBOR market model, based on freezing the drift in the forward rate dynamics.

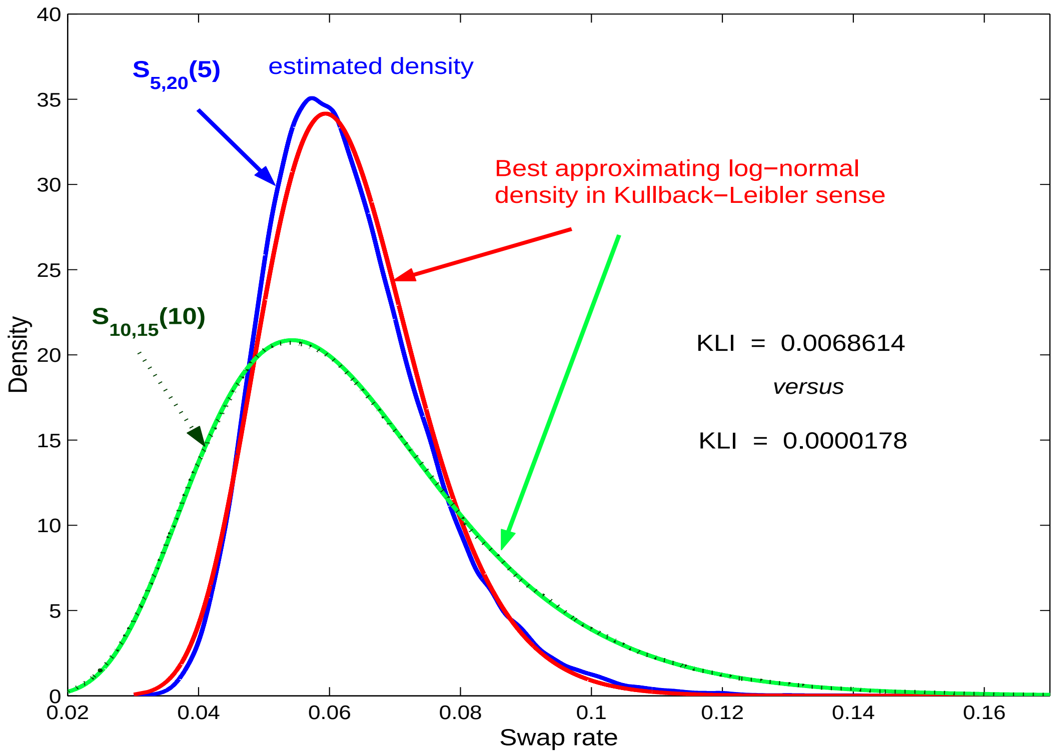

Figure 1: Estimated density functions of $S_{5,20}(5)$ (case (1.c)) and $S_{10,15}(10)$ (case (1.b)) compared to the best lognormal densities in KL sense.

References

[1] AMARI, S-I.(1985). *Differential Geometric Methods in Statistics*. Lecture Notes in Statistics, 28. Springer-Verlag, Berlin.

[2] BRACE, A., DUN, T., and BARTON, G. (1998). Towards a Central Interest Rate Model. *FMMA notes* working paper, published in *Handbooks in Mathematical Finance: Topics in Option Pricing, Interest Rates and Risk Management* (2001), Cambridge University Press

[3] BRACE, A., GATAREK, D., and MUSIELA, M. (1997). The market model of interest rate dynamics. *Mathematical Finance*, Vol. 7, 127–154.

[4] BRIGO, D. (1999). Diffusion Processes, Manifolds of Exponential Densities, and Nonlinear Filtering. in: O.E. Barndorff-Nielsen and E. B. Vedel Jensen (Editors), *Geometry in Present Day Science*, World Scientific.

[5] BRIGO, D., and HANZON, B. (1998). On some filtering problems arising in mathematical finance. *Insurance: Mathematics and Economics*, 22 (1) pp. 53-64.

[6] BRIGO, D., HANZON, B., and LE GLAND, F. (1999). Approximate Nonlinear Filtering by Projection on Exponential Manifolds of Densities. *Bernoulli*, Vol. 5, N. 3 (1999), pp. 495–534.

[7] D. BRIGO, and MERCURIO, F. (2001). *Interest Rate Models: Theory and Practice.* Spriger, Berlin.

[8] DUDEWICZ, E.J., and VAN DER MEULEN, E.C. (1987). The empiric entropy, a new approach to nonparametric entropy estimation, in Puri M.L. et al. eds., *New Perspectives in Theoretical and Applied Statistics*, Wiley, New York, 207-227

[9] IVANOV, A.V., and ROZHKOVA (1981). Properties of the statistical estimate of the entropy of a random vector with a probability density. *Problems of Information Transistion*, 17, 171-178.

[10] JAMSHIDIAN, F. (1997). Libor and swap market models and measures. *Finance and Stochastics*, 4, pp. 293–330.

[11] KAGAN, A.M. , LINNIK, Y.V., and RAO, C.R. (1973). *Characterization problems in Mathematical Statistics.* John Wiley and Sons, New York.

[12] KLÖDEN, P.E., and PLATEN, E. (1995) *Numerical Solutions of Stochastic Differential Equations.* Springer, Berlin.

[13] MATSUMOTO, K. (2001). Lognormal Swap Approximation in the LIBOR Market Model and its Application, *The Journal of Computational Finance* 5 107-131.

[14] MILLER, E. (2003). A New Class of Entropy Estimators for Multi-Dimensional Densities. *International Conference on Acoustics, Speech, and Signal Processing.* To appear.

[15] MILTERSEN K.R., SANDMANN K., and SONDERMANN D. (1997). Closed form solutions for term structure derivatives with log–normal interest rates. *Journal of Finance*, 52, pp. 409–430.

[16] RAPISARDA, F., BRIGO, D., and MERCURIO, F. (2002). Parameterizing Correlations: A Geometric Interpretation. Banca IMI Working Paper.

[17] REBONATO, R. (1998). *Interest Rate Option Models.* Second Edition. Wiley, Chichester, 1998.

[18] REBONATO, R. (2003). Modern pricing of interest rate derivatives: The LIBOR market model and beyond, Princeton University Press, forthcoming.

[19] SIDENIUS, J. (2000). LIBOR Market Models in Practice, *The Journal of Computational Finance* 3 5-26.

[20] VASICEK, O. (1976). A test for normality based on sample entropy. *J. R. Statist. Soc. B*, 38, pp. 54–59.