Dynamics of Implied Volatility Surfaces.

Article · February 2002

Source: CiteSeer

CITATIONS
101

READS
4,192

2 authors, including:

Rama Cont
University of Oxford
244 PUBLICATIONS **19,454** CITATIONS

SEE PROFILE

QUANTITATIVE FINANCE VOLUME 2 (2002) 45–60 RESEARCH PAPER
INSTITUTE OF PHYSICS PUBLISHING quant.iop.org

Dynamics of implied volatility surfaces

Rama Cont[1,3] **and José da Fonseca**[2]

[1] Centre de Mathématiques Appliquées, Ecole Polytechnique, F-91128 Palaiseau, France
[2] Ecole Superieure d'Ingenierie Leonard de Vinci, F-92916 Paris La Défense, France

E-mail: Rama.Cont@polytechnique.fr and jose.da_fonseca@devinci.fr

Received 20 September 2001
Published 4 February 2002
Online at stacks.iop.org/Quant/2/45

Abstract

The prices of index options at a given date are usually represented via the corresponding implied volatility surface, presenting skew/smile features and term structure which several models have attempted to reproduce. However, the implied volatility surface also changes dynamically over time in a way that is not taken into account by current modelling approaches, giving rise to 'Vega' risk in option portfolios. Using time series of option prices on the SP500 and FTSE indices, we study the deformation of this surface and show that it may be represented as a randomly fluctuating surface driven by a small number of orthogonal random factors. We identify and interpret the shape of each of these factors, study their dynamics and their correlation with the underlying index. Our approach is based on a Karhunen–Loève decomposition of the daily variations of implied volatilities obtained from market data. A simple factor model compatible with the empirical observations is proposed. We illustrate how this approach models and improves the well known 'sticky moneyness' rule used by option traders for updating implied volatilities. Our approach gives a justification for use of 'Vega's for measuring volatility risk and provides a decomposition of volatility risk as a sum of contributions from empirically identifiable factors.

1. Introduction

Although the Black–Scholes formula is very popular among market practitioners, when applied to ('vanilla') call and put options it is very often reduced to a means of quoting option prices in terms of another parameter, the *implied* volatility. The implied volatility $\sigma_t^{\mathrm{BS}}(K, T)$ of a call option with strike level K and maturity date T actually depends on K and T, in contradiction with the Black–Scholes model. The function $\sigma_t^{\mathrm{BS}} : (K, T) \to \sigma_t^{\mathrm{BS}}(K, T)$ which represents this dependence is called the *implied volatility surface* at date t. Specifying the implied volatility surface at a given date is therefore synonymous with specifying prices of all (vanilla) calls and puts at that date.

Two features of this surface have captured the attention of researchers in financial modelling. First, the non-flat instantaneous profile of the surface, whether it be a 'smile', 'skew' or the existence of a term structure, points out the insufficiency of the Black–Scholes model for matching a set of option prices at a given time instant. Second, the level of implied volatilities changes with time, deforming the shape of the implied volatility surface [11]. The evolution in time of this surface captures the evolution of prices in the options market.

The shortcomings of the Black–Scholes option pricing model when compared to empirical data from the options

[3] http://www.cmap.polytechnique.fr/~rama/

market have led to the development of a considerable literature on alternative option pricing models, in which the dynamics of the underlying asset is considered to be a nonlinear diffusion, a jump-diffusion process or a latent volatility model. These models attempt to explain the various empirical deviations from the Black–Scholes model by introducing additional degrees of freedom in the model such as a local volatility function, a stochastic diffusion coefficient, jump intensities, jump amplitudes etc. However, these additional parameters describe the *infinitesimal* stochastic evolution of the underlying asset while the market usually quotes options directly in terms of their market-implied volatilities which are global quantities. In order to see whether the model reproduces empirical observations, one has to relate these two representations: the infinitesimal description via a stochastic differential equation on one hand, and the market description via implied volatilities on the other hand.

However, in the majority of these models it is impossible to compute directly the shape of the implied volatility surface in terms of the model parameters. Although it is possible to compute the implied volatility surface numerically, these numerical studies show that simple jump processes and one factor stochastic volatility models do not reproduce correctly the profiles of empirically observed implied volatility surfaces and smiles [7, 10, 13, 40]. This problem is also reflected in the difficulty in 'calibrating' model parameters simultaneously to a set of liquid option prices on a given date: if the number of input option prices exceeds the number of parameters (which should be the case for a parsimonious model) a conflict arises between different calibration constraints since the implied volatility pattern predicted by the model does not correspond to the empirically observed one.

This problem, already present at the static level, becomes more acute if one examines the consistency of model dynamics with those observed in the options market. While a model with a large number of parameters—such as a non-parametric local volatility function or implied tree—may calibrate well the strike profile and term structure of options on a given day, the same model parameters might give a poor fit at the next date, creating the need for constant re-calibration of the model. Examples of such inconsistencies over time have been documented for the implied-tree approach by Dumas *et al* [17]. This time instability of model parameters leads to large variations in sensitivities and hedge parameters, which is problematic for risk management applications.

The inability of models based on the underlying asset to describe dynamic behaviour of option prices or their implied volatilities is not, however, simply due to the mis-specification of the underlying stochastic process. There is a deeper reason: since the creation of organized option markets in 1973, these markets have become increasingly autonomous and option prices are driven, in addition to movements in the underlying asset, also by internal supply and demand in the options market. This fact is also supported by recent empirical evidence of violations of qualitative dynamical relations between options and their underlying [6]. This observation can be accounted for by introducing sources of randomness which are specific to the options market and which are not present in the underlying asset dynamics [12].

In order to prevent the instability of model parameters resulting from this misspecification, it is necessary for the model to recognize these extra sources of randomness specific to the options market and incorporate the statistical features of their dynamics in the model. The aim of this work is to move a step forward in this direction and identify, through the study of dynamical features of implied volatility surfaces for two major option markets, some of the salient dynamic properties of option prices.

Based on an empirical study of time series of implied volatilities of SP500 and FTSE index options, we show that the implied volatility surface can be described as a randomly fluctuating surface driven by a small number of factors. Using a Karhunen–Loève decomposition, we recover and interpret the shape of these factors and study their dynamics, in relation to the underlying index. The implied volatility surface is then modelled as a stationary random field with a covariance structure matching the empirical observations. This model extends and improves the well known 'constant smile' or 'sticky delta' model [3] used by practitioners, while matching some salient features of implied volatility time series.

Our stochastic implied volatility model allows a simple description of the time evolution of a set of options and provides a rationale for Vega hedging of portfolios of options. This modelling approach also allows us to construct a Monte Carlo framework for simulating scenarios for the joint behaviour of a portfolio of call or put options, leading to a considerable gain in computation time for scenario generation.

Section 2 defines the implied volatility surface and the parametrization used. Section 3.2 describes the non-parametric smoothing procedure for constructing implied volatility surfaces from data. Section 3 describes the Karhunen–Loève decomposition and its implementation. This method is then applied to two data sets: SP500 index options (section 4) and FTSE index options (section 5). The stylized empirical features of these two data sets are summarized in section 6.1. Based on these empirical observations, a factor model for the implied volatility surface is proposed in section 6.2. Section 7 presents implications for risk management, indicates potential applications and discusses topics for further research.

2. Implied volatility surfaces

2.1. Definitions and notations

Recall that a European call option on a non-dividend paying asset S_t with maturity date T and strike price K is defined as a contingent claim with a payoff of $(S_T - K)^+$. Denoting by $\tau = T - t$ the time remaining to maturity, the Black–Scholes formula for the value of this call option with

$$C_{\mathrm{BS}}(S_t, K, \tau, \sigma) = S_t N(d_1) - K\mathrm{e}^{-r\tau} N(d_2) \tag{1}$$

$$d_1 = \frac{-\ln m + \tau(r + \frac{\sigma^2}{2})}{\sigma\sqrt{\tau}} \qquad d_2 = \frac{-\ln m + \tau(r - \frac{\sigma^2}{2})}{\sigma\sqrt{\tau}} \tag{2}$$

where $m = K/S_t$ is the moneyness and $N(u) = (2\pi)^{-1/2}\int_{-\infty}^{u}\exp(-\frac{z^2}{2})\,dz$. Let us now consider, in a market where the hypotheses of the Black–Scholes model do not necessarily hold, a call option whose (observed) market price is denoted by $C_t^*(K,T)$. The Black–Scholes implied volatility $\sigma_t^{BS}(K,T)$ of the option is then defined as the value of the volatility parameter which equates the market price with the price given by the Black–Scholes formula

$$\begin{aligned} &\exists!\quad \sigma_t^{BS}(K,T) > 0,\\ &C_{BS}(S_t,K,\tau,\sigma_t^{BS}(K,T)) = C_t^*(K,T). \end{aligned} \tag{3}$$

The existence and uniqueness of the implied volatility is due to the fact that the value of a call option as a function of volatility is a monotonic mapping from $[0,+\infty[$ to $]0,S_t - Ke^{-r\tau}[$. For fixed (K,T), $\sigma_t^{BS}(K,T)$ is in general a stochastic process and, for fixed t, its value depends on the characteristics of the option such as the maturity T and the strike level K: the function

$$\sigma_t^{BS} : (K,T) \to \sigma_t^{BS}(K,T) \tag{4}$$

is called the *implied volatility surface* at date t. Using the *moneyness* $m = K/S_t$ of the option, one can also represent the implied volatility surface in relative coordinates, as a function of moneyness and time to maturity: $I_t(m,\tau) = \sigma_t^{BS}(mS(t), t+\tau)$. This representation is convenient since there is usually a range of moneyness around $m = 1$ for which the options are liquid and therefore empirical data are most readily available.

2.2. Implied volatility as a state variable

As opposed to 'local' volatility, uniquely defined as the instantaneous increment of the quadratic variation of the underlying asset, there are a multitude of implied volatilities associated with a given asset, one for each traded option. Therefore, one is faced right from the start with a multivariate problem.

Also, implied volatilities represent prices of traded options on the same underlying asset, which are linked by various arbitrage relations. It is therefore obvious that implied volatilities for different strike and maturity levels cannot evolve independently and that they form a highly correlated multivariate system. Arbitrage restrictions on the dynamics of implied volatilities have been discussed by Schonbucher [38] in the case of a single implied volatility and more recently by Ledoit and Santa Clara [31] in the case of an implied volatility surface.

The above points seem to indicate that defining a model in terms of implied volatility rather than local volatility may complicate somewhat the modelling procedure. However, using implied volatility as state variable has some important advantages which greatly simplify the analysis.

First, implied volatilities are *observable*: they are directly derived from market data without making any modelling assumptions on the processes involved. By contrast, the local volatility or conditional variance of returns is not directly observable and has to be filtered out either from price data using a conditional volatility model (e.g. GARCH) or 'calibrated' with options data. In the first case, the quantity obtained is model dependent and in the second case it is the solution of a non-trivial and ill-posed inverse problem [8, 18].

Second, implied volatilities give a representation of the state of the option market which is familiar to the practitioner. A market scenario described in terms of the level or behaviour of implied volatility is easier to understand for a practitioner than the same statement rephrased, for instance, in terms of a local volatility or a jump intensity.

Third, shifts in the level of implied volatility are highly correlated across strikes and maturities, which suggests that their joint dynamics is driven by a small number of factors and makes parsimonious modelling of their joint dynamics possible.

Finally, implied volatility is increasingly used as a market reference by various market participants, as attested by the recent emergence of implied volatility indexes and derivative instruments associated with them, and also as a market risk indicator [32].

All these points motivate the study of the behaviour of implied volatility surfaces as a foundation for a market-based approach for modelling the dynamics of the options market.

2.3. Empirical study of implied volatility

At a static level, the dependence of implied volatility on strike and maturity has been studied in many markets by various authors. While the Black–Scholes model predicts a flat profile for the implied volatility surface $I_t(m,\tau)$, it is a well documented empirical fact that it exhibits both a non-flat strike and term structure [11, 13, 17, 21, 25, 27, 33]. A typical illustration is given in figure 1 in the case of SP500 index options.

The dynamic properties of implied volatility time series have also been studied by many authors in the literature; however most of these studies either focus on the term structure of *at the money* implied volatilities or study separately implied volatility smiles for different maturities. In other words, they study a cross-section of the surface in one direction, either moneyness or maturity, while maintaining the other parameter fixed. One then obtains a series of curves (smiles or term structures) to which most authors apply a principal component analysis (PCA).

The term structure of at the money implied volatilities was studied in [27, 29, 30, 41]. In particular, Avellaneda and Zhu [41] performed a PCA of the term structure of at-the-money implied volatilities, modelling them with a GARCH process. Härdle and Schmidt [29] perform a similar study on the term structure of the VDAX and report the presence of level, shift and curvature components in the deformation of the term structure. Das and Sundaram [13] study the term structure of implied skewness and kurtosis, showing that empirical patterns differ from those predicted by simple models.

Skiadopoulos *et al* [39] perform a PCA of implied volatility smiles of SP500 American options traded on the CME for different maturity buckets and distinguish two significant principal components. Alexander [1] performs a

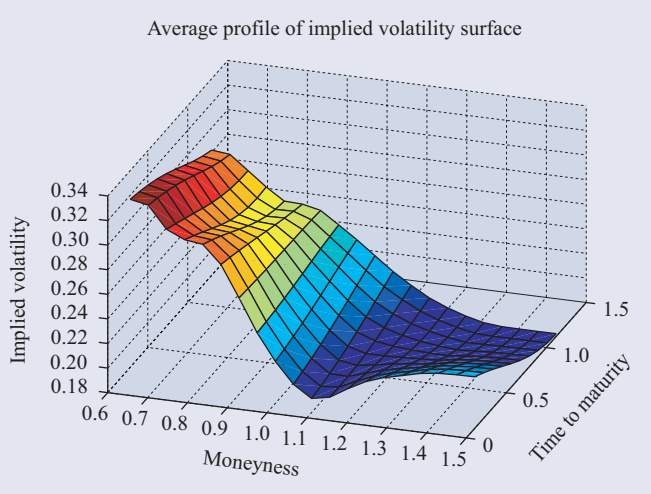

Figure 1. Typical profile of the implied volatility of SP500 options as a function of time to maturity and moneyness, March 1999.

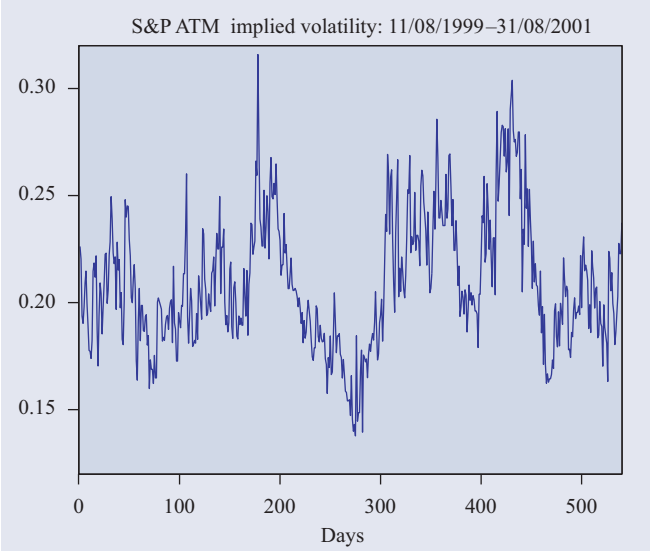

Figure 2. Evolution of at the money implied volatility for SP500 options, Aug. 1999–Aug. 2001.

similar analysis but using the deviation of implied volatilities from the at-the-money volatility. Hafner and Wallmeier [24] study the dynamics of smile curves for DAX options. Fengler *et al* [20], use a common principal components approach to perform a joint PCA on implied volatility smiles of different maturity buckets.

By contrast to the studies mentioned above, our aim is to consider, following our earlier work on SP500 options [11], the joint dynamics of all implied volatilities quoted on the market, looking simultaneously at all available maturity and moneyness values.

2.4. Deterministic models for implied volatility surfaces

The observation that the implied volatility surface evolves with time has led practitioners to develop simple rules to estimate its evolution [14]. A commonly used rule is the 'sticky moneyness' rule which stipulates that, when viewed in relative coordinates (m, τ), the surface $I_t(m, \tau)$ remains constant from day to day:

$$\forall (m, \tau), \qquad I_{t+\Delta t}(m, \tau) = I_t(m, \tau). \tag{5}$$

Another well known rule is the so-called 'sticky strike' rule: in this case one assumes that the level of implied volatilities for a given option (i.e. in absolute strike and maturity coordinates) does not change:

$$\forall (K, T), \qquad \sigma^{\mathrm{BS}}_{t+\Delta t}(K, T) = \sigma^{\mathrm{BS}}_t(K, T). \tag{6}$$

These rules are described in more detail in [14] and [33].

Before proceeding further, let us note that the above rules are actually deterministic laws of motion for the implied volatility surface: given today's option prices, there is no uncertainty (according to these rules) as to the value of implied volatilities tomorrow, up to errors due to smoothing/interpolation of today's implied volatility surface[4].

[4] More generally one can consider models where $I_t(\cdot, \cdot)$ or $\sigma^{\mathrm{BS}}_t(\cdot, \cdot)$ evolve in a deterministic way. In a detailed study Balland [3] shows that in fact the only arbitrage-free models in which $\sigma^{\mathrm{BS}}_t(\cdot, \cdot)$ is deterministic are Black–Scholes models with deterministic (that is, time dependent but not strike dependent) volatility. He also shows that if I_t is deterministic then the underlying risk neutral dynamics is a process with independent log-increments.

Unfortunately these simple rules are not verified in practice: as shown in figures 6 and 17, the implied volatility surface $I_t(m, \tau)$ has a noticeable standard deviation that cannot be neglected when considering either hedging or risk management of portfolios of options. Similarly, figure 2 shows the evolution of the at-the-money implied volatility for SP500 index options: as seen in this figure, even the at the money implied volatility, assumed to be constant or 'slowly varying' in the fixed smile model, fluctuates between 15 and 40% in less than a few months. Insufficiency of this 'fixed smile' approach has also been pointed out in previous studies [32, 35].

In this empirical study we precisely focus on these fluctuations of the surface $I_t(m, \tau)$: what is the nature of these fluctuations? How can they be quantified, modelled in a parsimonious fashion and simulated in a simple way?

3. Methodology

We first describe in section 3.1 the data sets used. From the data we construct a time series of smooth surfaces, via a procedure described in section 3.2. This time series of smooth surfaces is then modelled as a stationary random surface to which we apply a Karhunen–Loève decomposition: section 3.3 presents the method and section 3.4 discusses how it is applied to the analysis of implied volatility surfaces.

3.1. Data sets

The data sets studied contain end-of-day prices of European-style call and put options on two major indices: the SP500 index and the FTSE 100 index. We observe every day implied volatilities for traded options $I_t(m_i, \tau_i)$, $i = 1 \ldots n$ where n is

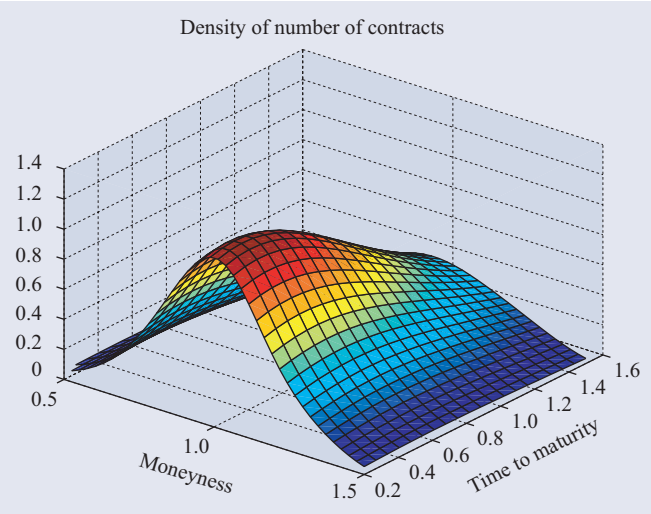

Figure 3. Distribution of observations with respect to moneyness and time to maturity: SP500 options.

the number of options actively traded on that day. Typically n is around 100. Time to maturity τ range between a month and a year. Moneyness values outside the interval [0.5, 1.5] are filtered out since the numerical uncertainty on implied volatility may be too high and the liquidity very low. All options used are out of the money options: calls are used for $m > 1$, puts for $m < 1$. These are precisely the options which contain the most information about implied volatility movements.

Implied volatilities are observed for a discrete set of moneyness and time to maturity values; the grid of observation is both irregular and changing with time as the level of the underlying fluctuates. Figure 3 shows the distribution of the number of observations as a function of the level of moneyness and time to maturity. As illustrated in this figure, the number of traded strikes decreases as we move far from the money and as time to maturity increases. Also, the maximum number of observations occurs at the money ($m = 1$) and not forward-at-the-money ($m = -r\tau$) which would create a linear bias in the position of the peak as τ increases. Also, the number of observations does not seem to be skewed towards $m > 1$ or $m < 1$ (this would advocate for example the use of log-moneyness instead of moneyness). Hence the choice of m as coordinate seems reasonable.

Similar properties are observed for FTSE options (figure 15).

3.2. Construction of smooth volatility surfaces

In order to obtain an implied volatility for arbitrary strikes and maturities, the practice is to interpolate or smooth the discrete data. This can be done either with a parametric form or in a non-parametric way. For instance, it is common practice in many banks to use (piecewise) polynomial functions to fit the implied volatility smile [17]. These choices are driven more by convenience than by any fundamental consideration. Given the absence of arguments pointing to a specific parametric form, the approach we have used here is a non-parametric one: the filtered data set is used to construct for each day a smooth estimator of the implied volatility surface, defined on a fixed grid, using a non-parametric Nadaraya–Watson estimator [2, 28]:

$$\hat{I}_t(m,\tau) = \frac{\sum_{i=1}^{n} I_t(m_i,\tau_i)g(m-m_i,\tau-\tau_i)}{\sum_{i=1}^{n} g(m-m_i,\tau-\tau_i)} \qquad (7)$$

where $g(x, y) = (2\pi)^{-1}\exp(-x^2/2h_1)\exp(-y^2/2h_2)$ is a Gaussian kernel. I_t is simply the convolution of the observed data points with the smoothing kernel g. The choice of the Gaussian kernel instead of another kernel does not influence the results very much. The important parameters are the bandwidth parameters h_1, h_2, which determine the degree of smoothing. Too small values will lead to a bumpy surface, too large ones will smooth away important details. The bandwidth can be determined using a cross-validation criterion [28]. A more efficient method is to use an adaptive bandwidth estimator in order to obtain an 'optimal' bandwidth h [9, 23]. Large sample properties of these estimators have been extensively studied, we refer to the monograph by Härdle [28] or to [2].

We thus obtain a daily time series $(I_t(m,\tau), t = 0 \dots N)$ of smooth implied volatility surfaces $I : [m_{\min}, m_{\max}] \times [\tau_{\min}, \tau_{\max}] \to [0, \infty[$ to which we apply an analysis of variance, described in the next section. Note that $\tau_{\min} > 0$: it is around 2 weeks. In particular, we do not use points very close to maturity since it is well known that, in this range, implied volatilities may have highly irregular behaviour which makes them unexploitable for risk management purposes. Also, we do not attempt to extrapolate to $\tau = 0$ since such an extrapolation will highly depend on the (arbitrary) choice of the estimator.

3.3. Principal component analysis of random surfaces

In order to analyse the joint dynamics of all implied volatilities, we model the implied volatility surface as a stationary random field, to which we apply a Karhunen–Loève decomposition. The Karhunen–Loève decomposition is a generalization of PCA to higher dimensional random fields. Let $A = [m_{\min}, m_{\max}] \times [\tau_{\min}, \tau_{\max}]$ be the range of values of moneyness and time-to-maturity.

A surface parametrized by $A \subset R^2$ is a smooth map $u : A \to R$. For two surfaces u, v define the scalar product $\langle u, v\rangle$ by

$$\langle u, v\rangle = \int_A u(x)v(x)\,\mathrm{d}x \qquad (8)$$

where $x = (m, \tau)$ designates a point in the moneyness/maturity plane. Let $U(\omega, x), x \in A$ be a two-dimensional real-valued random field, defined as a family of random variables indexed by a two-dimensional parameter $x \in A$, such that each realization of the random field U is a smooth surface $U(\omega, \cdot) : A \to R$. One can then view the random field $U(\omega, x)$ as a random surface $U(\omega, \cdot)$.

Viewing U as a family of real-valued random variables, one may define the covariance coefficients of two points on the surface as

$$K(x_1, x_2) = \mathrm{cov}(U(x_1), U(x_2)) \qquad x_1, x_2 \in A. \qquad (9)$$

The function $K(x_1, x_2), x_1, x_2 \in A$ is analogous to the covariance matrix of a random vector. One may also define the correlation coefficients $C(x_1, x_2)$. The covariance coefficients define the kernel of an integral operator $\mathcal{K}$ acting on a surface u through

$$\mathcal{K}u(x) = \int_A K(x, y)u(y)\,\mathrm{d}y. \tag{10}$$

$\mathcal{K}$ is a bounded symmetric positive operator; its eigenvectors f_n define an orthonormal family (f_n). Let $v_1^2 \geqslant v_2^2 \geqslant \cdots \geqslant 0$ the associated eigenvalues: $\mathcal{K} \cdot f_n = v_n^2 f_n$. Each eigenfunction f_n is therefore a surface $f_n : A \subset R^2 \to R$ which is the solution of a Fredholm integral equation defined by the kernel K:

$$\int K(x, y) f_n(y)\,\mathrm{d}y = v_n^2 f_n(x). \tag{11}$$

The singular value decomposition of the covariance operator $\mathcal{K}$ can translate into a decomposition of the kernel K:

$$K(x_1, x_2) = \sum v_n^2 f_n(x_1) f_n(x_2). \tag{12}$$

Let $U_n(\omega, x) = \langle U(\omega, \cdot), f_n \rangle$ be the (random) projection of the random surface U on the eigenfunction f_n. By orthogonality of the (f_n), (U_n) is a sequence of uncorrelated random variables and

$$U(\omega, \cdot) = \sum U_n(\omega) f_n(\cdot) = \sum U_n f_n. \tag{13}$$

Equation (13), which expresses the random field U as a superposition of uncorrelated random variables $(U_n \cdot f_n)$, each of which is the product of a scalar random variable with a deterministic surface, is called the Karhunen–Loève decomposition of the random surface U. Note that the random variables $(U_n \cdot f_n)$ are orthogonal both as random variables—elements of $L^2(\Omega, P)$—*and* as surfaces—elements of $L^2(A)$—which makes the Karhunen–Loève decomposition very convenient for computational purposes.

3.4. Karhunen–Loève decomposition: numerical implementation

Given the high autocorrelation, skewness and positivity constraints on the implied volatility itself, we apply the above approach to the daily variations of the logarithm of implied volatility $\Delta X_t(m, \tau) = \ln I_t(m, \tau) - \ln I_{t-1}(m, \tau)$. First, we use market implied volatilities to produce a smooth implied volatility surface $I_t(m, \tau)$ using the non-parametric procedure described in section 3.2. Next, we compute the daily variation of the logarithm of implied volatility for each point on the surface: $\Delta X_t(m, \tau) = \ln I_t(m, \tau) - \ln I_{t-1}(m, \tau)$. We then apply a Karhunen–Loève decomposition to the random field $U_t(m, \tau) = \Delta X_t(m, \tau)$. To this end, we compute the sample covariance coefficients $K(m, \tau, m', \tau') = \mathrm{cov}(U_t(m, \tau), U_t(m', \tau'))$, which specify the covariance operator $\mathcal{K}$.

Formally, we are faced with an eigenvalue problem for an operator in a function space. In order to solve the problem numerically, we reduce it to a finite-dimensional one using a Galerkin procedure for the Fredholm equation (11), as described below. Choosing a family of smooth functions (surfaces) (h_n), one expands each eigenfunction on this basis:

$$f_i(m, \tau) = \sum_{n=1}^{N} a_{ij} h_j(m, \tau) + \epsilon_N. \tag{14}$$

For example, one can choose as approximation functions (h_j) spline functions, commonly used for interpolating volatility surfaces and yield curves in financial applications. Substituting the truncated sum into equation (11) yields an error term

$$\varepsilon_N = \sum_{n=1}^{N} a_{ij} \left(\int_A \mathrm{d}m'\,\mathrm{d}\tau' K(m, m', \tau, \tau') h_j(m', \tau') - v_j^2 h_j(m, \tau) \right).$$

The Galerkin method consists in requiring that the error ε_N be orthogonal to the approximating functions $(h_n, n = 1 \ldots N)$:

$$\forall j = 1 \ldots N, \quad \langle \varepsilon_N, h_j \rangle = 0$$

$$\iff \sum_{n=1}^{N} a_{ij} \left[\int_A \mathrm{d}x\, h_i(x) \int_A \mathrm{d}x' K(x, x') h_j(x') - v_j^2 \int_A \mathrm{d}x\, h_i(x) h_j(x) \right] = 0.$$

Using matrix notation

$$A = [a_{ij}] \tag{15}$$

$$B_{ij} = \langle h_i, h_j \rangle \tag{16}$$

$$C_{ij} = \int_A \mathrm{d}m\,\mathrm{d}\tau \int_A \mathrm{d}m'\,\mathrm{d}\tau' h_i(m, \tau) K(m, m', \tau, \tau') h_j(m', \tau') \tag{17}$$

$$D = \mathrm{diag}(v_i^2, i = 1 \ldots N) \tag{18}$$

the orthogonality condition (15) can be rewritten

$$CA = DBA \tag{19}$$

where C and B are symmetric positive matrices, computed from the data. Numerically solving the generalized eigenvalue problem (19) for D and A and substituting the coefficients of A in (14) yields the eigenmodes f_k. D yields their associated variances v_k^2. We then compute the projection $x_k(t)$ of $X_t(m, \tau)$ on the eigenmode f_k:

$$x_k(t) = \langle X_t - X_0, f_k \rangle = \int_A X_t(m, \tau) f_k(m, \tau)\,\mathrm{d}m\,\mathrm{d}\tau \tag{20}$$

and analyse the scalar time series $(x_k(t), t = 1 \ldots N)$ (principal component processes). The surface then has the representation

$$I_t(m, \tau) = I_0(m, \tau) \exp\left(\sum_k x_k(t) f_k \right). \tag{21}$$

Finally, we analyse the correlation between the movements in implied volatilities and the underlying asset returns by computing the correlation coefficients between $\Delta x_k(t) = x_k(t) - x_k(t-1)$ and the underlying returns $\ln S(t) - \ln S(t-1)$ for each principal component k.

We now describe the results obtained by applying this technique to two sets of data: SP500 index options (section 4) and FTSE index options (section 5).

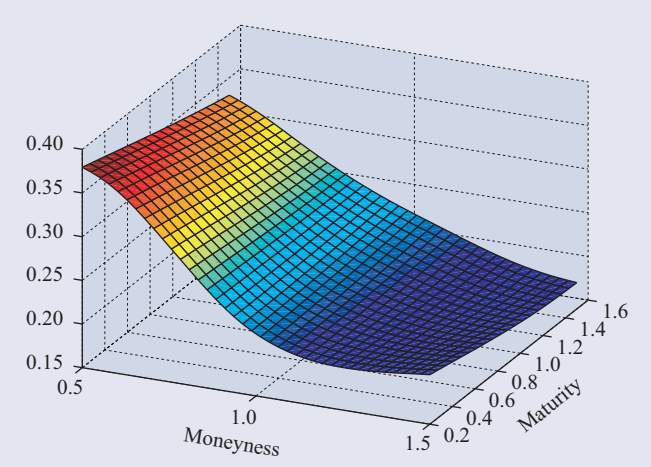

Figure 4. Average implied volatility surface for SP500 options.

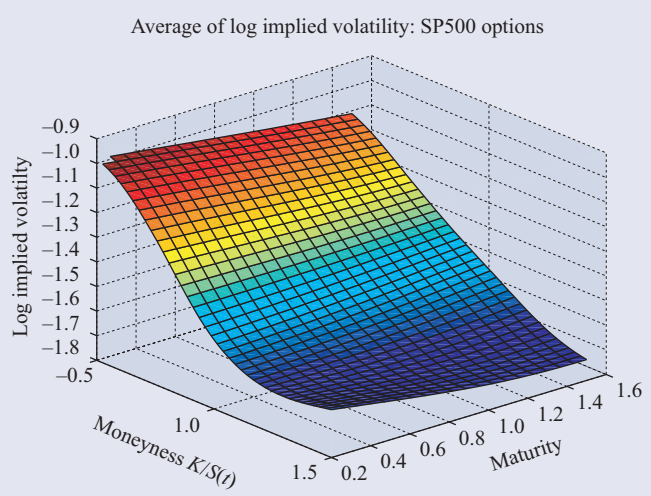

Figure 5. Average log-implied volatility surface, SP500 options.

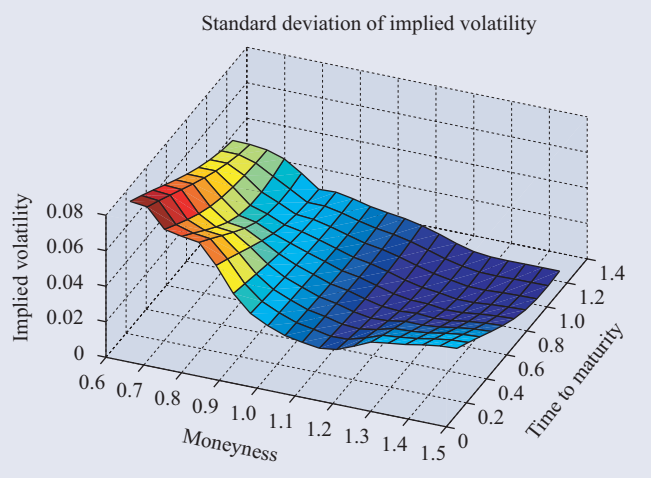

Figure 6. Daily standard deviation of SP500 implied volatilities as a function of moneyness and time to maturity.

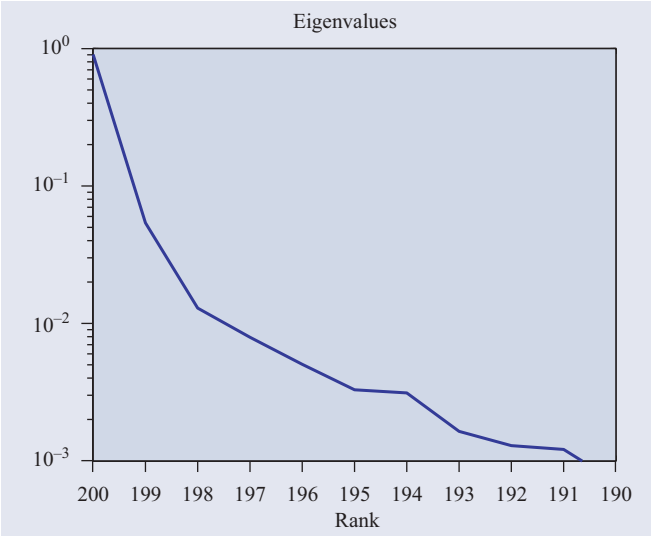

Figure 7. Sorted eigenvalues of covariance operator as a function of their rank for daily variations of SP500 implied volatilities.

4. Empirical results for SP500 index options

The first data set contains end of day prices for all traded European style call and put options on the SP500 index between 2 March 2000 and 2 February 2001. These options are traded on the Chicago Board of Options Exchange and differ from the American style options, studied in [39], traded on the Chicago Mercantile Exchange.

Using the time series of smooth volatility surfaces estimated via the procedure described in section 3.2, we compute the sample average of the random volatility surface $I_t(m, \tau)$, shown in figure 4. Here the average implied volatility is shown as a function of time to maturity in years and moneyness $m = K/S(t)$. The figure shows a decreasing profile in moneyness ('skew') as well as a downward sloping term structure.

Figure 5 shows the surface obtained when logarithms of implied volatilities are taken before averaging. Note that, while the skew persists, the term structure is less pronounced. This remark also pleads in favour of studying logarithms of implied volatilities instead of volatilities themselves.

The sample standard deviation of implied volatilities, shown in figure 6, illustrates that the surface is not static and fluctuates around its average profile. This figure can also be seen as a test of the the 'fixed smile' model [3], which would predict a small or negligible variability in implied volatilities, plotted in moneyness coordinates. Clearly this is far from being the case. More precisely, the daily standard deviation of the implied volatility can be as large as a third of its typical value for out of the money options, resulting in an important impact on option prices.

The eigenvalues ν_k^2 decay quickly with k (figure 7), showing that a low-dimensional factor model gives a good approximation for the dynamics of the surface. In fact the first three eigenmodes account for 98% of the variance, so in what follows we shall focus on their properties only.

The shape of the first eigenmode is shown in figure 8. All of its components are positive: a positive shock in the

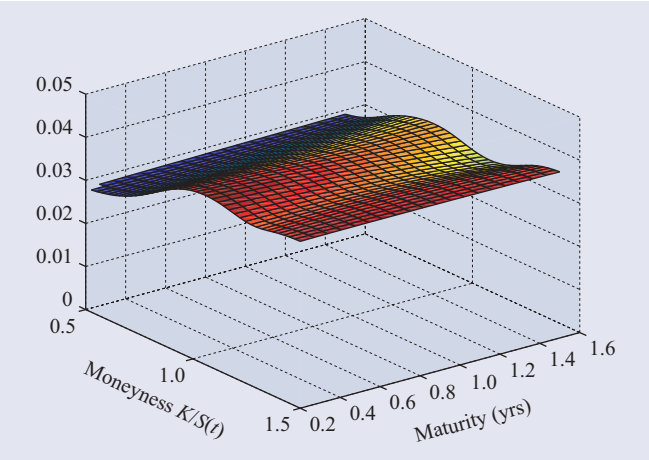

Figure 8. First eigenmode of daily implied volatility variations for SP500 index options. This eigenmode, which accounts for around 80% of the daily variance of implied volatilities, can be interpreted as a level effect.

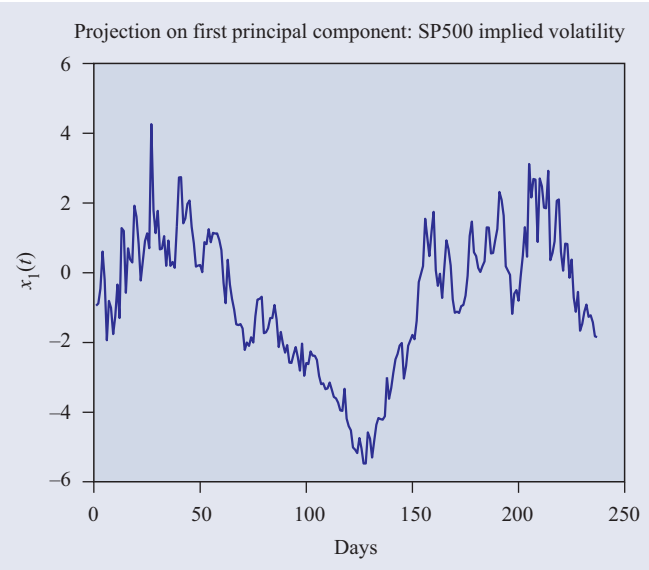

Figure 9. Time evolution of the projection $x_1(t)$ of SP500 implied volatilities on the first principal component.

direction of this eigenmode results in a global increase of all the implied volatilities: it may therefore be interpreted as a 'level' factor. Figure 9 shows the projection $x_1(t)$ of the surface on this eigenmode: it is observed to have a mean-reverting behaviour with a typical mean-reversion time of around a month. Figure 10 shows the autocorrelation function of the principal component process $x_1(t)$: the autocorrelation coefficients are significantly positive up to a month, showing persistence in the values of the principal component process. As the partial autocorrelation coefficients show, this persistence can be conveniently represented by a low-order autoregressive process: an AR(1)/Ornstein–Uhlenbeck process already captures most of the persistence. Estimating an AR(1) process on the series gives an autoregression constant of 0.965 which gives a mean reversion time of 28 days. This time scale should be compared with the maturity of liquid options which is

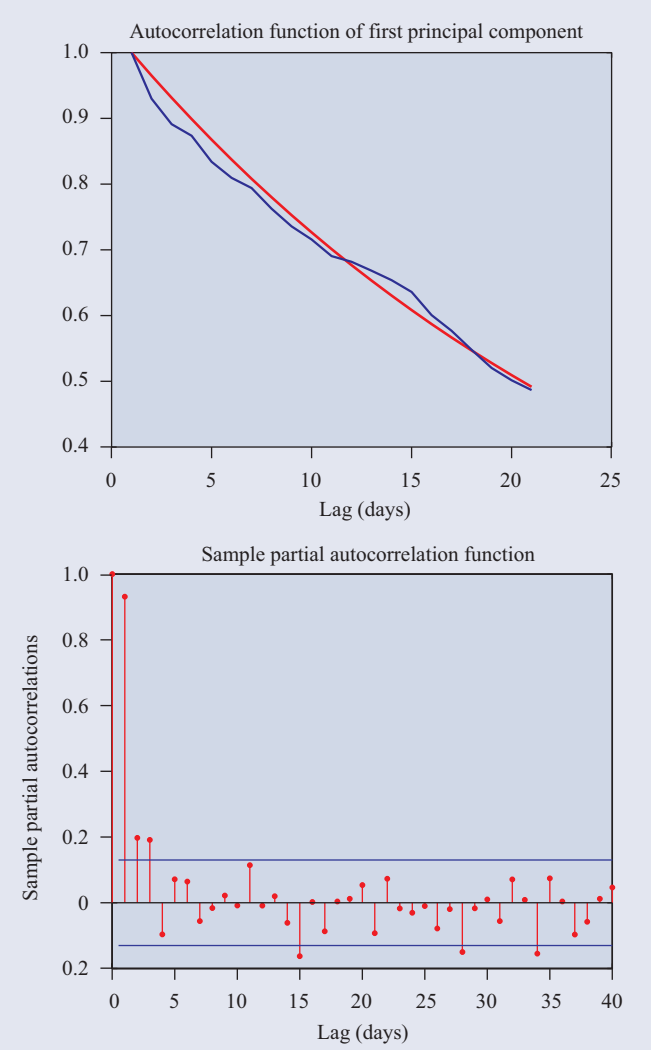

Figure 10. Top: in blue, autocorrelation function of $x_1(t)$; in red, decaying exponential with time constant of 28 days. Bottom: partial autocorrelation coefficients. Time unit: days.

around 3 months: they are both of the same magnitude showing that mean reversion of implied volatility is neither 'slow' nor 'fast' but happens at the time scale of the option maturity.

Movements along this eigenmode, which account for around 80% of the daily variance, have a strong *negative* correlation with the underlying index returns. This observation is consistent with the so-called leverage effect: the overall negative correlation of implied volatility and asset returns. Note also that although the dependence in τ is weak, it is not zero, and multiplying the surface by a constant can amplify this dependence. Since the variance is large along this direction, large movements along this direction can induce term structure fluctuations.

The second eigenmode, in contrast, changes sign at the money: the components are positive for $m > 1$ and negative for $m < 1$ (figure 11). A positive shock along this direction increases the volatilities of out of the money calls while decreasing those of out of the money *puts*. By biasing the implied volatilities towards high strikes, positive movements along this direction increase the skewness of the risk neutral

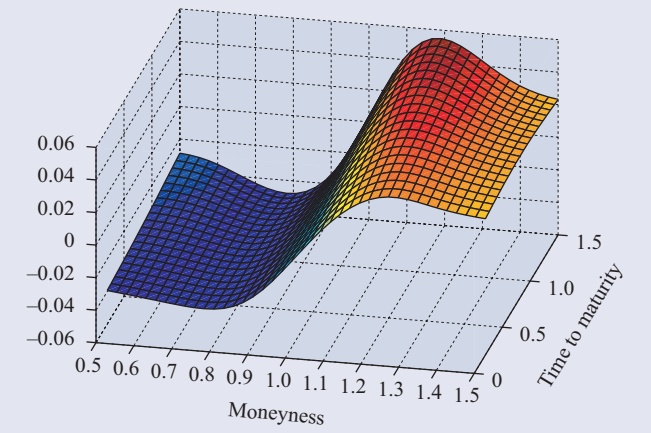

Figure 11. Second eigenmode of daily implied volatility variations for SP500 index options.

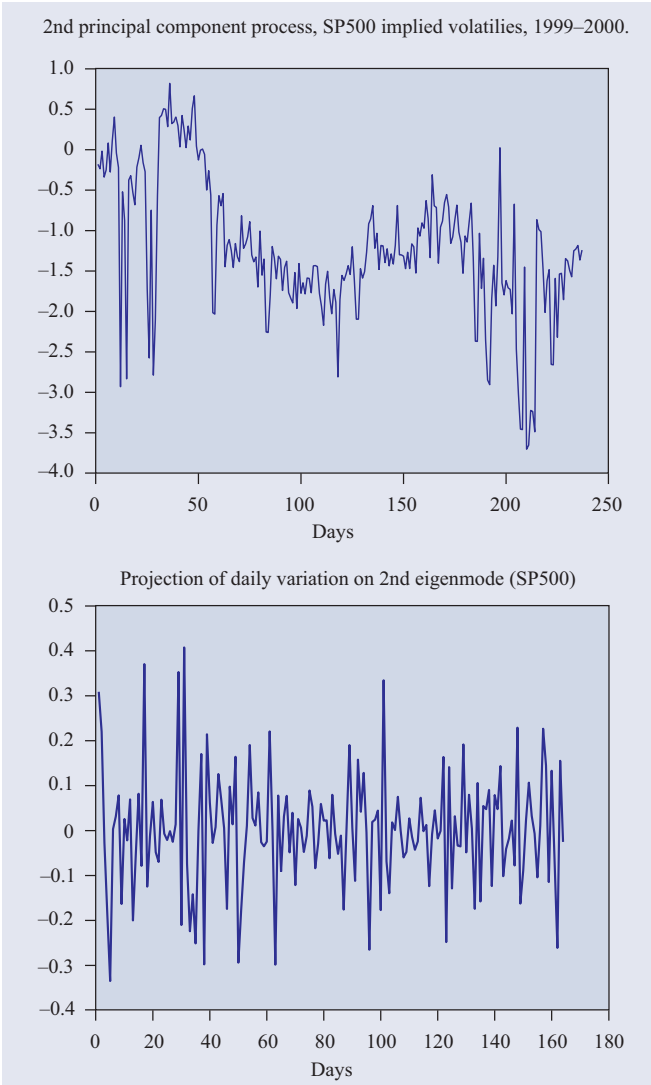

Figure 12. Top: time evolution of the projection $x_2(t)$ of SP500 implied volatilities on the second principal component. Bottom: daily increments $\Delta x_2(t)$. Time unit: days.

density. Note also that this mode exhibits a systematic slope with respect to maturity, which accounts for the term structure observed above[5] (figure 6). Figure 12 shows the evolution of the projection of the surface on this eigenmode. Again, it exhibits a mean-reverting behaviour but with more intermittency, showing jumps and spikes. Inspection of the partial autocorrelation function, shown in figure 13, shows that the correlation structure can be well approximated by an AR(1) process. Estimation of an AR(1) model gives an autoregression coefficient of 0.924 which corresponds to a mean reversion timescale of 12.6 days.

In contrast to what was observed for the first eigenmode, movements along this eigenmode are only weakly correlated with the underlying returns. This suggests that the leverage effect is not a general feature of implied volatility movements but is related to movements in the general level of implied volatilities, while relative movements of implied volatilities can be decorrelated with respect to the underlying. For a fixed maturity, the cross sections of this mode correspond to the shape found by Skiadopoulos *et al* [39] on another data set (SP500 *futures* options traded on the CME).

The third eigenmode, shown in figure 14, is a 'butterfly' mode which is interpreted as a change in the convexity of the surface, accompanied by a downward sloping term structure which biases the magnitude of the fluctuations towards short term implied volatilities. In terms of risk, movements along this direction contribute to the 'fattening' of both the tails of the risk neutral density. This component only contributes to 0.8% of the overall variance.

Table 1 gives summary statistics for the time series of the first three principal component processes. Each of them is highly autocorrelated and mean-reverting with a mean reversion time close to a month: figures 9 and 12 illustrate the sample paths. Inspection of the partial autocorrelation functions show that the autocorrelation structures (shown in figures 10 and 13) are well approximated by AR(1)/Ornstein–Uhlenbeck processes. Although the unconditional distributions are not Gaussian (all series have excess kurtosis and some skewness), the deviation from normality is mild. Similar results were found for DAX options in [24]. Also, none of these factors is perfectly correlated with the underlying returns. The largest (negative) correlation value is between the underlying and the first factor, corresponding to a value of -66%, but still remains significantly smaller than 1. Movements in other directions, which reflect changes in the shape of the surface, seem to have little correlation with the underlying asset.

5. Empirical results for FTSE options

In this section we present the results obtained when applying the methods described in section 3 to a database of FTSE 100 index options. The options under study are European call and put options on the FTSE 100 index (ESX options). The database contains more than 2 years of daily transaction prices and quoted Black–Scholes implied volatilities for all quoted

[5] Actually figure 11 shows a *downward* sloping term structure but recall that eigenmodes are defined up to a sign and normalization so the sign is irrelevant.

Table 1. Summary statistics for principal component times series: SP500 index options.

Eigenmode	Daily standard deviation	Proportion of variance (%)	Kurtosis (1 day)	Skewness (1 day)	Mean reversion time (days)	Correlation with underlying
1	0.10	94	6.4	0.4	28	−0.66
2	0.09	3	7.9	0.15	12.6	≃0
3	0.05	0.8	7.8	0	22	0.27

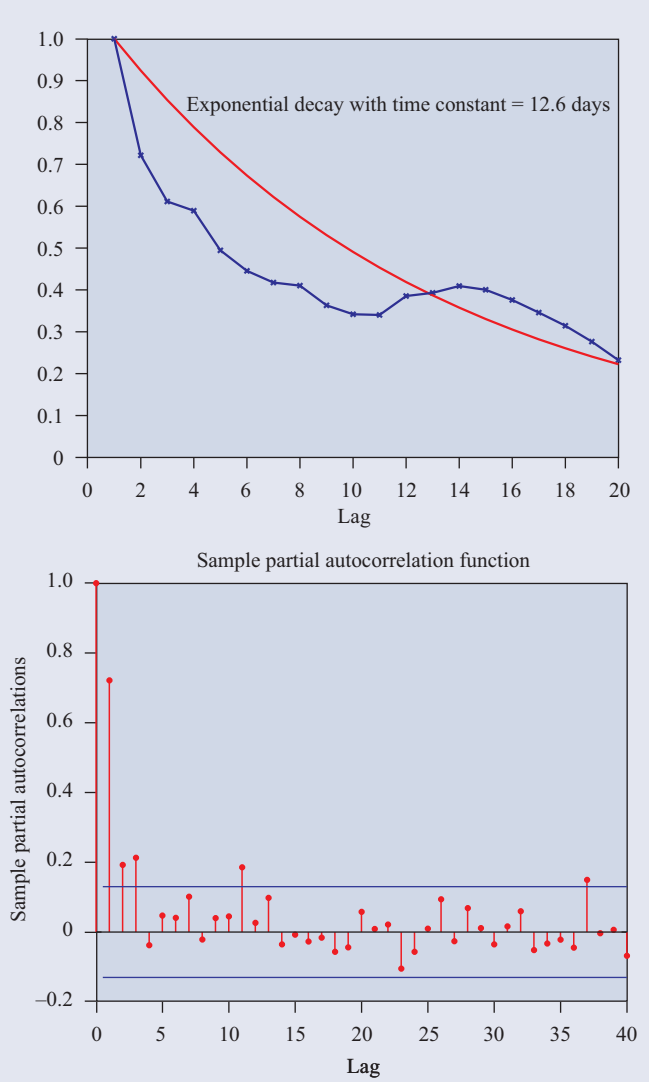

Figure 13. 2nd principal component process, SP500 implied volatilities. Top: autocorrelation function of $x_2(t)$ (blue) compared to exponential decay (red) with a time constant of 12.6 days. Bottom: partial autocorrelation coefficients of $x_2(t)$.

options on the LIFFE[6] market (Aug.–Sept. 2001). Maturities range from one month to more than a year, moneyness levels from 50 to 150% with respect to the underlying. Observations are unevenly distributed in time-to-maturity and moneyness. Figure 15 shows the frequency of observations for a given moneyness and time to maturity as a function of these two variables, obtained via a non-parametric kernel estimator; it has a shape similar to the one observed for SP500 options (see

[6] London International Financial Futures Exchange.

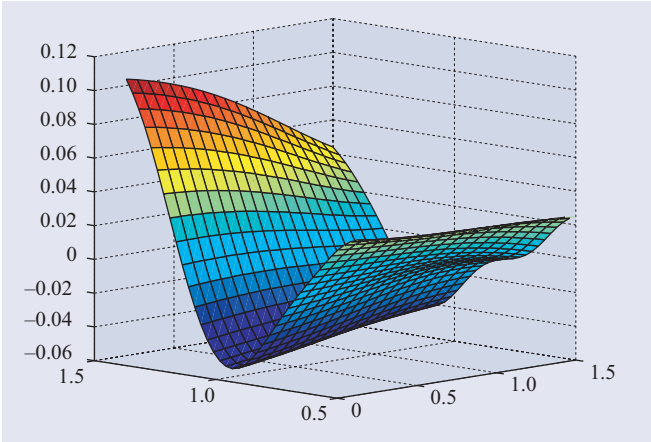

Figure 14. Third eigenmode of daily implied volatility variations for SP500 index options.

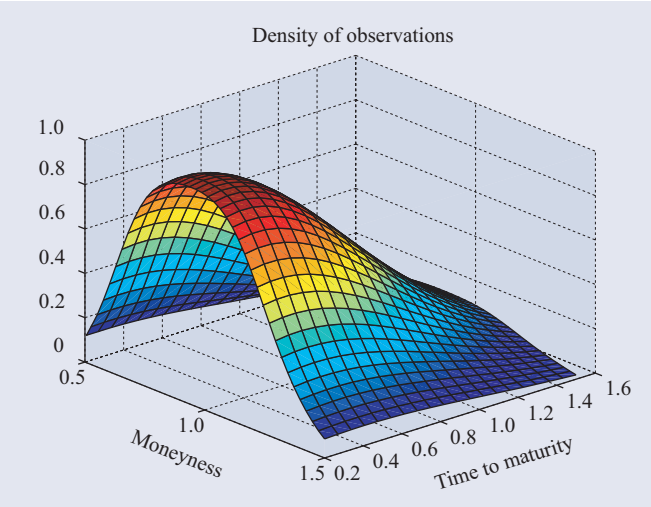

Figure 15. Distribution of observations with respect to moneyness and time to maturity: FTSE options.

section 3.1).

As in the preceding section, we first construct a time series of smooth surfaces by using the non-parametric smoothing procedure used in section 3.2. We then perform an analysis of some of the dynamical properties of the smoothed time series.

We first compute the sample mean of the surface. The average profile of the implied volatility surface is shown in figure 16, on a logarithmic scale. As in the case of the SP500 index options, it displays a strong skew, the implied volatility decreasing as a function of moneyness, as well as some curvature in the moneyness direction. Note that the term structure is relatively flat. However, as seen from the standard

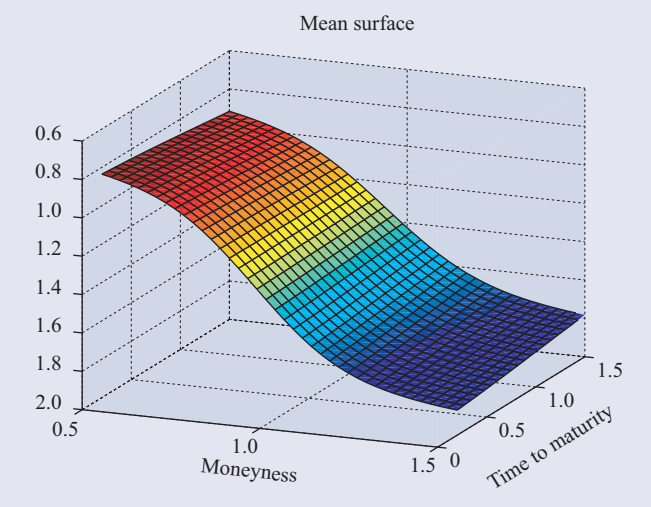

Figure 16. Average profile of FTSE implied volatility surface.

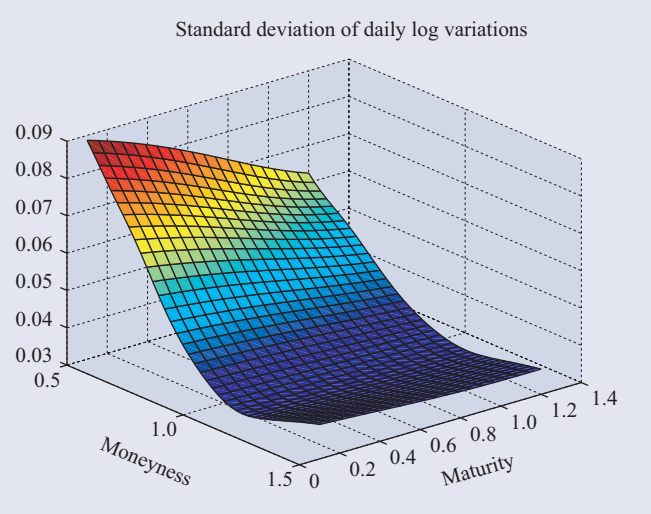

Figure 17. Standard deviation of daily variations in FTSE implied volatilities.

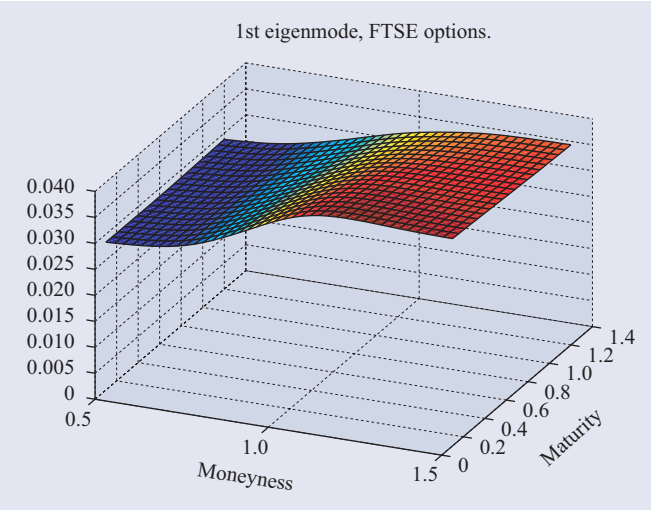

Figure 18. First eigenmode of daily implied volatility variations for FTSE index options.

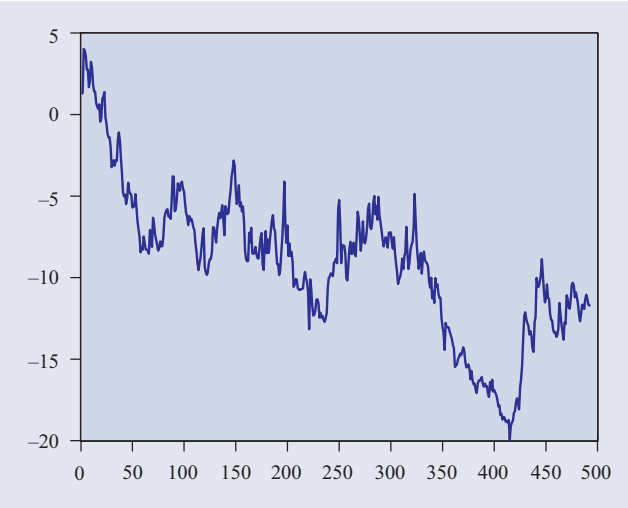

Figure 19. Projection $x_1(t)$ on first eigenmode, FTSE options.

deviation of their daily log-increments (figure 17), the daily *variations* of implied volatilities do display a systematic term structure: short term implied volatilities fluctuate more than longer ones.

Applying the Karhunen–Loève decomposition (section 3) to the daily log-variations of the implied volatility yields the eigenmodes (principal components) and their associated eigenvalues (variances). We then project the implied volatility surface each day on each eigenmode and study the principal component time series thus obtained.

The variances, ranked in decreasing order, decrease very quickly: as in the case of SP500 index options, the variance of daily variations is well captured by the first three principal components which account for more than 95% of the observed daily variations.

The shape of the dominant eigenmode is shown in figure 18. This eigenmode contains around 96% of the daily variance. Similarly to the first eigenmode for the SP500 surface, all of its components are positive: a positive shock in the direction of this eigenmode results in a global increase of all the implied volatilities: it may therefore be interpreted as a 'level' factor. The projection of the implied volatility surface on this eigenmode is shown in figure 19: this time series is highly persistent mean-reverting. The one day lagged partial autocorrelation coefficient is significant and falls to insignificant levels after 1 day, pointing to an AR(1) correlation structure. Estimating an AR(1) model on the data gives an autocorrelation coefficient of 0.9808 which corresponds to a mean reversion time of 51 days, reflecting a high degree of persistence. The autocorrelation function, shown in figure 20, is well represented by the corresponding exponential decay over a timescale of 51 days. As in the case of the SP500 options, the mean reversion time is close to the lifetime of an option, not much shorter nor longer.

The second eigenmode, shown in figure 21, reveals a clear slope effect: the signs of the coordinates are negative for moneyness levels $m < 1$ (out of the money puts) and positive for $m > 1$ (out of the money calls), changing sign at the money.

Table 2. Statistics for principal component times series: FTSE options.

Eigenmode	Daily standard deviation	Proportion of variance (%)	Kurtosis of Δx_k (1 day)	Mean reversion time (days)	Correlation with underlying
1	2.9	96	9.3	51	0.7
2	0.43	2	5.7	65	0.08
3	0.27	0.8	7.5	81	0.7

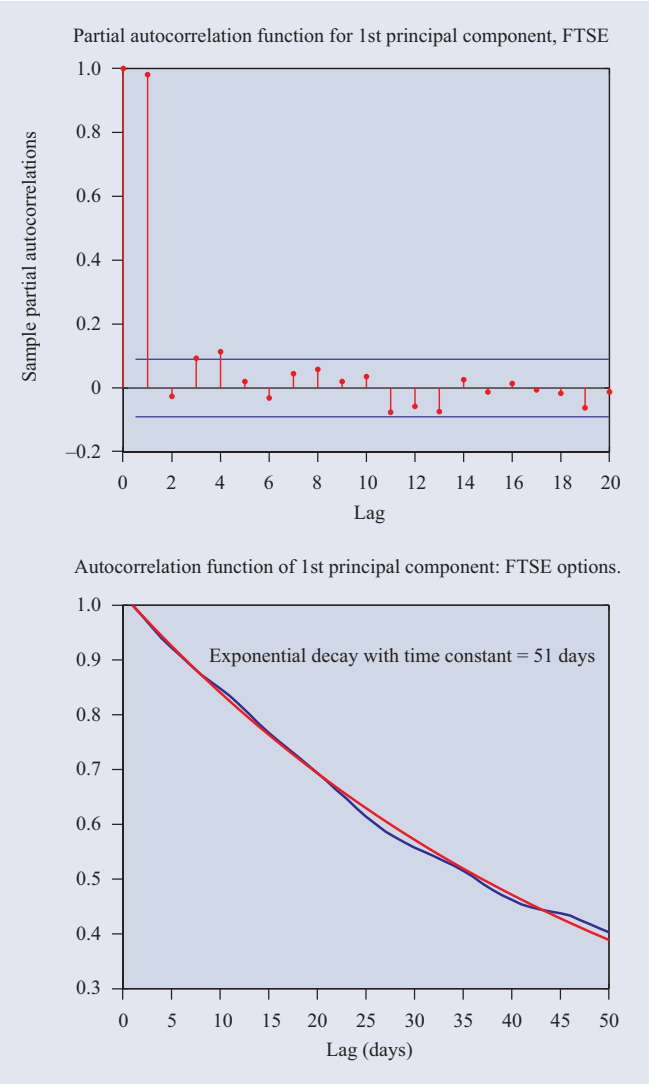

Figure 20. Top: partial autocorrelation coefficients of the projection of FTSE implied volatilities on the first principal component. Bottom: autocorrelation function of $x_1(t)$ compared to an exponential decay with time constant = 51 days. Time unit: days.

A positive shock in this direction therefore results in a decrease in the prices of out of the money puts accompanied by a rise in the prices of out of the money calls. In terms of the risk neutral distribution it corresponds to a (positive) increase in the skewness.

The third eigenmode, shown in figure 24, again represents changes in the curvature/convexity of the surface in the strike direction, associated with a smaller slope effect in the maturity direction.

Table 2 gives summary statistics for the time series of the

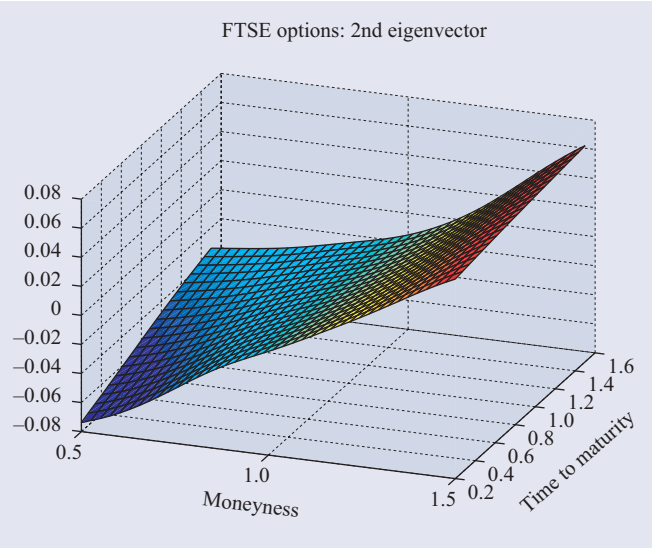

Figure 21. Second eigenmode of daily implied volatility variations for FTSE index options.

first three factor loadings. As in the case of SP500 options, they exhibit persistence and mean reversion. Their correlation structure is well approximated by an AR(1) process with mean reversion times around two months. However, for the second and third principal components both the unconditional distributions and the distribution of residuals have significant non-Gaussian features and exhibit intermittency, suggesting the use of an AR(1) model with non-Gaussian noise.

6. A mean reverting factor model for the implied volatility surface

6.1. Summary of empirical observations

Let us now summarize the statistical properties of implied volatility surfaces observed in these two data sets:

(1) Implied volatilities display high (positive) autocorrelation and mean reverting behaviour.
(2) The variance of the daily log-variations in implied volatility can be satisfactorily explained in terms of a small number of principal components (two or three).
(3) The first eigenmode reflects an overall shift in the level of all implied volatilities.
(4) The second eigenmode reflects opposite movements in (out of the money) call and put implied volatilities.
(5) The third eigenmode reflects changes in the convexity of the surface.

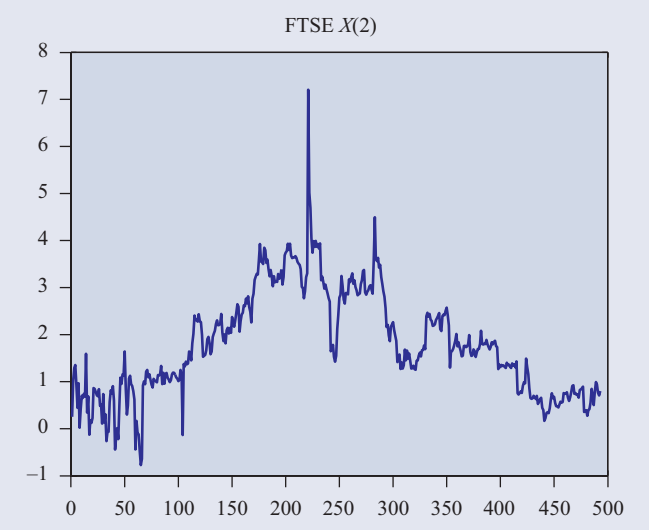

Figure 22. Projection $x_2(t)$ on second eigenmode, FTSE options.

(6) Movements in implied volatility are not perfectly correlated with movements in the underlying asset.
(7) Shifts in the global level of implied volatilities are negatively correlated with the returns of the underlying asset.
(8) Relative movements of implied volatilities have little correlation with the underlying.
(9) The projections of the surface on its eigenmodes (principal component processes) exhibit high (positive) autocorrelation and mean reversion over a timescale close to a month.
(10) The autocorrelation structure of principal component processes is well approximated by the AR(1)/Ornstein–Uhlenbeck process.

Similar results have also been obtained for DAX options, omitted here for the sake of brevity.

6.2. A factor model for the implied volatility surface

Based on these empirical results, we propose a flexible class of factor models which is compatible with these observations: in our framework the implied volatility surface is directly used as the state variable and modelled as a stationary random surface evolving in a low-dimensional manifold of surfaces. The (log-) implied volatility surface $X_t(m,\tau)$ is represented by the sum of the initial surface and its fluctuations along the principal directions $(f_k, k = 1 \dots d)$:

$$\ln I_t(m,\tau) = X_t(m,\tau) = X_0(m,\tau) + \sum_{k=1}^{d} x_k(t)\, f_k(m,\tau) \tag{22}$$

$$\langle f_j, f_k \rangle = \delta_{j,k} \qquad x_k(t) = \langle X_t - X_0, f_k \rangle \tag{23}$$

where $X_0(m,\tau) = \ln I_0(m,\tau)$ is a constant surface and the components x_k are Ornstein–Uhlenbeck processes driven by independent noise sources Z_k, which can be Wiener or jump processes:

$$\mathrm{d}x_k(t) = -\lambda_k(x_k(t) - \overline{x_k})\,\mathrm{d}t + \gamma_k\,\mathrm{d}Z_k(t). \tag{24}$$

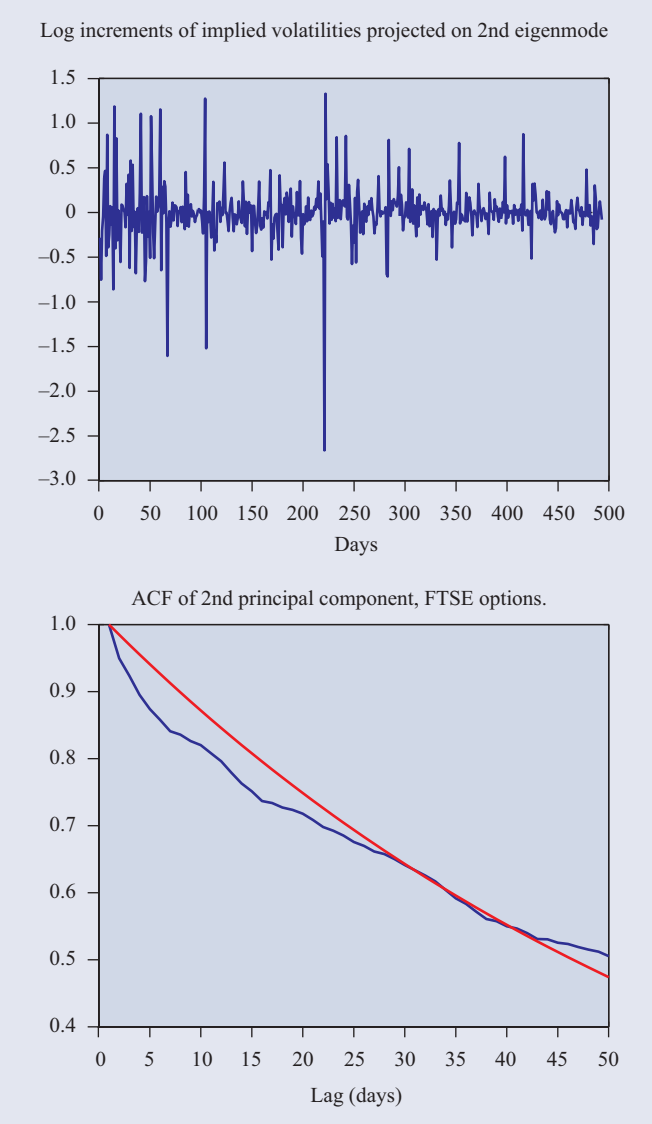

Figure 23. Top: time evolution of the projection of FTSE implied volatilities on the second principal component. Bottom: autocorrelation function of $x_2(t)$ (blue) compared with exponential decay with a time constant of 51 days (red). Time unit: days.

Here λ_k represents the speed of mean reversion along the kth eigenmode, $\overline{x_k}$ the long-term average value and γ_k is the volatility of implied vols along this direction. The 'constant smile' or 'sticky delta' model corresponds to the case where $I_t = I_0$, i.e. where $x_k(t) = 0$. Our factor model is therefore an extension of the sticky delta model.

In the stationary case, these parameters are simply related to the empirical observations: denoting by ν_k^2 the variance of the kth principal component, then

$$\overline{x_k} = \langle \overline{X} - X_0, f_k \rangle \tag{25}$$

$$\nu_k^2 = \frac{\gamma_k^2}{2\lambda_k} \tag{26}$$

$$\mathrm{corr}(x_k(t), x_k(t+s)) = \exp(-\lambda_k s) \tag{27}$$

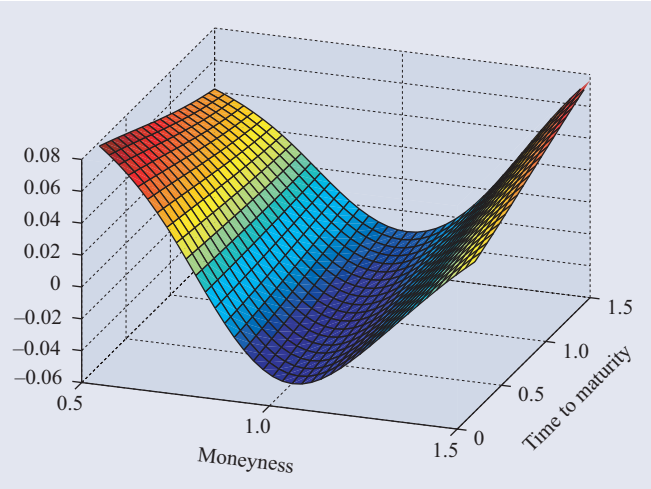

Figure 24. Third eigenmode of daily implied volatility variations for FTSE index options.

Table 3. Estimation results for factor model.

Data set	λ_1	λ_2	λ_3
SP500	0.035	0.08	0.045
FTSE	0.0193	0.0152	0.0122

and may be estimated from discrete observations $x_k(t), t = 0, 1, 2 \ldots$ of the principal component processes: indeed, under the above hypothesis x_k follows an AR(1) process:

$$x_k(t+1) = \mathrm{e}^{-\lambda_k} x_k(t) + (1 - \mathrm{e}^{-\lambda_k})\overline{x_k} + \sigma_k \epsilon^k(t) \tag{28}$$

$$\epsilon(t) \text{ IID noise } \sigma_k = \gamma_k \sqrt{\frac{1 - \exp(-\lambda_k)}{2\lambda_k}}. \tag{29}$$

Estimation of the coefficients was done using a general method of moments procedure using the discrete observations $x_k(t), t = 1, 2, \ldots, T$. Estimation results are shown in table 3. In the case where Z_k are independent Wiener processes, the implied volatilities have a log-normal distribution, which both ensures positivity and is consistent with many empirical observations in the literature.

6.3. Dynamics of option prices

An interesting feature of using implied volatilities as state variables is that the price of any call option $C_t(T, K)$ is simply given by the Black–Scholes formula

$$\begin{aligned} C_t(K, T) &= C_{\mathrm{BS}}(S_t, K, \tau, \sigma_t^{\mathrm{BS}}(K, T)) \\ &= C_{\mathrm{BS}}\left(S_t, K, \tau, I_t\left(\frac{K}{S_t}, T - t\right)\right) \end{aligned} \tag{30}$$

and the continuous-time dynamics of $C_t(K, T)$ may be computed by applying Ito's lemma to equation (30) [12]. This allows to compute Vega hedge ratios and generate scenarios for call/put options directly using the factor model (22). This should be contrasted for example with the approach of Derman and Kani [15] where the *local* volatility surface is modelled as a random surface: in this case the impact on option prices is complicated and non-explicit.

Arbitrage constraints for this type of model were first discussed in a risk-neutral framework by Schönbucher [38] (see also [31]). However, in the case where the state variable is the whole volatility surface some care must be taken to define properly the available instruments and strategies in order to remain in the realm of applicability [12]. In contrast here we have studied the *historical* dynamics of call or put options: of course, this is the relevant point of view from a risk management perspective. Our approach bridges the gap between risk-neutral approaches [15, 31, 38] and the empirical work on historical dynamics of implied volatility [1,20,24,29,39,41] and allows automatic adjustment to today's option prices which are simply calibrated to the initial condition $I_0(m, \tau)$: $X_0(m, \tau) = \ln I_0(m, \tau)$. The link between the historical and risk neutral approaches is discussed in more detail in [12].

7. Applications and extensions

We have presented in this work an empirical analysis of historical co-movements of implied volatilities of options on the SP500 and FTSE indices. Using a Karhunen–Loève decomposition, we have shown that these fluctuations can be accounted through a low-dimensional, but not one-dimensional, multifactor model. We have extracted and interpreted the shapes of the principal component surfaces and showed that the principal component processes exhibit a mean reverting autoregressive behaviour similar to that of an Ornstein–Uhlenbeck process.

Of course the behaviour of these time series can be studied in more detail, going beyond second-order statistics and examining effects obtained by conditioning on various variables. However, already at this level, there are several interesting implications and potential applications, some of which we enumerate below.

7.1. The nature of volatility risk

An interesting feature of our data is that there appears to be more than one factor present in the movements of implied volatilities and these factors are not perfectly correlated with the underlying asset. While this may appear as no surprise to operators in the options market that implied volatilities of different strikes do not vary in unison, it is clearly not a feature of one-factor 'complete' market models which attempt to explain all movements of option prices in terms of the underlying asset. In other words: 'Vega' risk cannot be reduced to 'Delta' risk.

The presence of more than one factor also shows that bivariate stochastic volatility models may lack enough degrees of freedom to explain co-movements of implied volatilities other than general shifts in their level.

7.2. Relation with stochastic volatility models

Any stochastic model for instantaneous volatility also implies a model for the deformation of the implied volatility surface. This correspondence may be, however, very non-explicit [22, 34]. It would nevertheless be interesting to compare our

empirical findings with the dynamics of implied volatility surfaces in 'traditional' stochastic volatility models [4, 22], at least on a qualitative or numerical basis.

7.3. Quantifying and hedging volatility risk

Our model allows a simple and straightforward approach to the modelling and hedging of volatility risk, defined in terms familiar to practitioners in the options market, namely that of Vega risk defined via Black–Scholes Vegas. It also suggests a simple approach to the simulation of scenarios for the joint evolution of a portfolio of call and put options along with their underlying asset. Such approaches have already been considered in the case where the at-the-money implied vol is perturbated by random shocks while the shape of the smile is fixed: Malz [32] uses this approach for stress-testing; see also Rosenberg [35]. Our framework allows an extension to the case of a random smile/surface. These points shall be studied in a forthcoming work [12].

7.4. Generating scenarios for option portfolios

Given a stochastic model for the joint behaviour of the underlying and its implied volatility surface, one can in principle generate scenarios and compute confidence intervals for future values of the implied volatility. Given the monotonic relation between implied volatility and option prices, this can allow us to compute confidence intervals for the values of various call and put options and portfolios composed of these instruments. This may be an interesting approach for computing confidence intervals for future scenarios in the option market, but needs a more detailed characterization of the link between the underlying asset dynamics and that of the option market, perhaps going beyond covariance measures as we have done above.

7.5. Dynamics of volatility indices

Another field in which this approach has potentially interesting applications is the hedging and risk management of 'volatility derivatives' such as options on an implied volatility index. There are several such market indices at present: VIX (CBOE), VDAX (Frankfurt) and VXN (Nasdaq) are some examples. Such an index, if well diversified across strikes and maturities, will be sensitive to the factors described above (mainly the first factor).

In summary, we find that the study of dynamics of implied volatility surfaces points to the possibility of a simple dynamic representation which seems to have interesting applications and links with other topics in financial econometrics and option pricing theory. We hope to pursue some of these research topics in the near future.

Acknowledgments

We thank Carol Alexander, Marco Avellaneda, Raphael Douady, Nicole El Karoui, Vincent Lacoste, Ole Barndorff Nielsen, Eric Reiner and seminar participants at Aarhus University, CNRS Aussois Winter School, Nikkei Symposium on Empirical Finance, HEC Montreal, Bedlewo workshop on Mathematical Finance, the Verona Conference on Credit and Market Risk, the AMS-SMF 2001 Congress (Lyon, France) and RISK Math Week 2000 (London and New York) for their comments and suggestions.

References

[1] Alexander C 2001 Principles of the skew *Risk* January S29–32
[2] Ait-Sahalia Y and Lo A 2000 Nonparametric estimation of state-price densities implicit in financial asset prices *J. Finance* **53** 499–548
[3] Balland P 2002 Deterministic implied volatility surfaces *Quant. Finance* **2** 31–44
[4] Barndorff-Nielsen O E and Shephard N 2001 Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion) *J. R. Stat. Soc.* B **63** 167–241
[5] Black F and Scholes M 1973 The pricing of options and corporate liabilities *J. Political Economy* **81** 637–54
[6] Bakshi G, Cao C and Chen Z 2000 Do call prices and the underlying stock always move in the same direction? *Rev. Financial Studies* **13** 549–84
[7] Bakshi G, Cao C and Chen Z 1997 Empirical performance of alternative option pricing models *J. Finance* **LII** 2003–49
[8] Berestycki H, Busca J and Florent I 2000 An inverse parabolic problem arising in finance *C. R. Acad. Sci. Paris I* **331** 965
[9] Brockmann M, Gasser T and Herrmann E 1993 Locally adaptive bandwidth choice for kernel regression estimators *J. Am. Stat. Assoc.* **88** 1302–9
[10] Busca J and Cont R 2002 Implied volatility surfaces in exponential Lévy models *Working Paper*
[11] Cont R and da Fonseca J 2001 Deformation of implied volatility surfaces: an empirical analysis *Empirical Approaches to Financial Fluctuations* ed Takayasu (Tokyo: Springer)
[12] Cont R, da Fonseca J and Durrleman V 2002 Stochastic models of implied volatility surfaces *Economic Notes* at press
[13] Das S R and Sundaram R K 1999 Of smiles and smirks: a term structure perspective *J. Financial Quant. Anal.* **34** 211–40
[14] Derman E 1998 Regimes of volatility *Risk*
[15] Derman E and Kani I 1998 Stochastic strike and term structure *Int. J. Theor. Appl. Finance* **1** 61–110
[16] Derman E, Kani I and Zou J Z 1995 The local volatility surface: unlocking the information in option prices *Goldman Sachs Quantitative Strategies Research Notes*
[17] Dumas B, Fleming J and Whaley R E 1998 Implied volatility functions: empirical tests *J. Finance* **8** 2059–106
[18] Dupire B 1993 Pricing with a smile *Risk* **7** 18–20
[19] Engle R and Rosenberg J 2000 Testing the volatility term structure using option hedging criteria *J. Derivatives* **8** 10–29
[20] Fengler M, Härdle W and Villa C 2000 The dynamics of implied volatilities: a common principal component approach, *Discussion paper* no 38/2001, SfB 373, Humboldt University, Germany
[21] Franks J R and Schwartz E J 1991 The stochastic behaviour of market variance implied in the price of index options *Econ. J.* **101** 1460–75
[22] Fouque J P, Papanicolaou G and Sircar R 2000 *Derivatives in Markets with Stochastic Volatility* (Cambridge: Cambridge University Press)
[23] Gasser T, Kneip A and Khler W 1991 A flexible and fast method for automatic smoothing *J. Am. Stat. Assoc.* **86** 643–52
[24] Hafner R and Wallmeier M 2000 The dynamics of DAX implied volatilities *Working Paper*

[25] Heynen R 1993 An empirical investigation of observed smile patterns *Review Futures Markets* **13** 317–53
[26] Hodges H M 1996 Arbitrage bounds on the implied volatility strike and term structures of European style options *J. Derivatives* **3** 23–35
[27] Heynen R, Kemma A and Vorst T 1994 Analysis of the term strucure of implied volatilities *J. Financial Quant. Anal.* **29** 31–56
[28] Härdle W 1990 *Applied Nonparametric Regression* (Cambridge: Cambridge University Press)
[29] Härdle W and Schmidt P 2000 Common factors governing VDAX movements and the maximum loss *Humboldt University Working Paper*
[30] Heynen R 1995 *Essays on Derivative Pricing Theory* (Amsterdam: Thesis Publishers)
[31] Ledoit O and Santa Clara P 1999 Relative pricing of options with stochastic volatility *UCLA Working Paper*
[32] Malz A 2001 Do implied volatilities provide an early warning of market stress? *J. Risk* **3** no 2
[33] Rebonato R 1999 *Volatility and Correlation in the Pricing of Equity, FX and Interest Rate Options* (New York: Wiley)
[34] Renault E and Touzi N 1996 Option hedging and implied volatilities in a stochastic volatility model *Math. Finance* **6** 279–302
[35] Rosenberg J V 2000 Implied volatility functions: a reprise *J. Derivatives* **7**
[36] Rubinstein M 1994 Implied binomial trees *J. Finance* **49** 771–818
[37] Sabbatini M and Linton O 1998 A GARCH model of the implied volatility of the Swiss market index from option prices *Int. J. Forecasting* **14** 199–213
[38] Schönbucher P J 1999 A market model for stochastic implied volatility *Phil. Trans. R. Soc.* A **357** 2071–92
[39] Skiadopoulos G, Hodges S and Clelow L 2000 Dynamics of the S&P500 implied volatility surface *Rev. Derivatives Res.* **3** 263–82
[40] Tompkins R 2001 Stock index futures markets: stochastic volatility models and smiles *J. Futures Markets* **21** 4378
[41] Zhu Y and Avellaneda M 1997 An E-ARCH model for the term structure of implied volatility of FX options *Appl. Math. Finance* **4** 81–100