SPECTRAL METHODS FOR VOLATILITY DERIVATIVES

Albanese, Claudio and Mijatovic, Aleksandar

Independent Consultant

1 March 2006

SPECTRAL METHODS FOR VOLATILITY DERIVATIVES

CLAUDIO ALBANESE, HARRY LO, AND ALEKSANDAR MIJATOVIĆ

Abstract. In the first quarter of 2006 Chicago Board Options Exchange (CBOE) introduced, as one of the listed products, options on its implied volatility index (VIX). This opened the challenge of developing a pricing framework that can simultaneously handle European options, forward-starts, options on the realized variance and options on the VIX. In this paper we propose a new approach to this problem using spectral methods. We define a stochastic volatility model with jumps and local volatility, which is almost stationary, and calibrate it to the European options on the S&P 500 for a broad range of strikes and maturities. We then extend the model, by lifting the corresponding Markov generator, to keep track of relevant path information, namely the realized variance. The lifted generator is too large a matrix to be diagonalized numerically. We overcome this difficulty by developing a new semi-analytic algorithm for block-diagonalization. This method enables us to evaluate numerically the joint distribution between the underlying stock price and the realized variance which in turn gives us a way of pricing consistently the European options, general accrued variance payoffs as well as forward-starts and VIX options.

1. Introduction

In recent years there has been much interest in trading derivative products whose underlying is a realized variance of some liquid financial instrument (e.g. S&P 500) over the life of the contract. The most popular payoff functions[1] are linear, leading to variance swaps, square root, yielding volatility swaps, and the usual put and call payoffs defining variance swaptions.

It is clear that the plethora of possible derivatives on the realized variance is closely related to the standard volatility-sensitive instruments like vanilla options, which are also exposed to other market risks, and the forward starting options which are almost pure vega bets and are mainly exposed to the movements of the forward smile. Recently Chicago Board Options Exchange (CBOE) introduced options on the volatility index[2] (VIX) which are also important predictors for the future behaviour of implied volatility. The main purpose of this paper is to introduce a framework in which all of the above financial instruments (i.e. the derivatives on the realized variance as well as the instruments depending on the implied volatility) can be priced and hedged consistently and efficiently.

Our central idea is very simple and can be described as follows. We define a dynamics for the underlying that includes local volatility, stochastic volatility and jumps and can be calibrated to the implied volatility surface for a wide variety of strikes and maturities (for the case of European options on the S&P 500 see figure 1). The underlying process is stationary as can be inferred from the fact that the implied forward volatility smile behaves in a consistent way (see figures 7, 8, 9 and 10). This is a consequence of the minimal explicit time-dependence in the calibration of the model.

There are two features of this model that make it possible to obtain the distributions of the future behaviour of implied volatility and of the realized variance of the underlying. The first feature is the complete numerical solubility of the model. In other words spectral theory provides a simple and efficient algorithm (see subsection 3.5) for obtaining a conditional probability distribution function for the underlying between any given pair of times in the future. This property is sufficient to determine completely the forward volatility smile and the distribution

[1]For the precise definition of these products see subsections 2.1 and 2.2.

[2]For a brief description of the these securities see subsection 2.3. For the definition of VIX see (CBOE 2003).

of VIX for any maturity. The second feature of the framework that makes it possible to deal with the realized variance is the extendability property of Markov generators known as *lifting* (for details see section 5). This allows us to define a new Markov generator of an extended process which keeps track both of the realized variance and the underlying forward rate. Using a block-diagonalization algorithm, described in section 6, and the standard methods from spectral theory we find a joint probability distribution function for the underlying process and its realized variance or volatility at any time in the future (for time horizons of 6 months, 1 year and 2 years see figures 13, 14 and 15 respectively). This joint pdf is precisely what is needed to price a completely general payoff which depends on the realized variance and on the underlying.

There are two natural and useful consequences of this approach. One is that we do not need to specify exogenously the process for the variance and then try to find an arbitrage-free dynamics for the underlying, but instead imply such a process from the observed vanilla market via the model for the underlying (a term-structure of the fair values of variance swaps as implied by the vanilla market data and our model is shown in figure 19). The second consequence is that this approach bypasses the use of Monte Carlo techniques and therefore yields sharp and easily computable sensitivities to the market parameters. This is because the pricing algorithm yields, as a by-product, all the necessary information for finding the required hedge-ratios.

There is a rapidly growing interest in trading volatility derivatives in financial markets which is mainly a consequence of the following two factors. On one hand pure volatility instruments are used to hedge implicit vega exposure of the portfolios of market participants, thus bypassing the need to trade frequently in the vanilla options market, which in itself is advantageous because of the relatively large bid-offer spreads prevailing in that market. On the other hand volatility derivatives are a useful tool for speculating on the future volatility levels and for trading the difference between realized and implied volatility.

This interest is reflected in the vast amount of literature devoted to volatility products. The analysis of the realized variance is intrinsically easier than that of realized volatility because of the additivity of the former. Under the hypothesis that the underlying price process is continuous the realized variance can be hedged perfectly by a European contract with the logarithmic payoff, first studied in (Neuberger 1994), and a dynamic trading strategy in the underlying. This approach does not require an explicit specification for the instantaneous volatility process of the underlying and can therefore be used within any stochastic volatility framework. This idea has been developed in (Carr & Madan 1998) and (Demeterfi, Derman, Kamal & Zou 1999*a*) where the static replication strategy for the log contract, using calls and puts, is described. Types of mark-to-market risk faced by a holder of a variance swap are studied and classified in (Chriss & Morokoff October 1999). A direct delta-hedging approach for the realized variance is given in (Heston & Nandi November 2000).

A shortcoming of pricing variance swaps without specifying a volatility model (as described in (Carr & Madan 1998) and (Demeterfi, Derman, Kamal & Zou 1999*b*)) is that this methodology does not yield a natural method for the computation of the sensitivities to market parameters (i.e. Greeks). In (Howison, Rafailidis & Rasmussen 2004) a diffusion model for the volatility process is specified which allowed the authors to use PDE technology to price and hedge variance and volatility swaps as well as more general payoffs. The obstacle here is that, even if one manages to guess the correct dynamics for the instantaneous volatility, the stochastic volatility models are known to have difficulties fitting the observed market skews for both the short dated and the long dated options at the same time.

The derivatives on the realized volatility can be considered naturally as the derivatives on the square root of the realized variance. In (Brockhaus & Long 1999) the authors provide a volatility convexity correction which relates the two families of derivatives. A practical difficulty with hedging a volatility swap using variance swaps is that it requires a dynamic position in the log contract which in turn depends on a strip of vanilla options. Some of these options will be very far out-of-the-money and therefore trading at large bid-offer spreads which would make the re-balancing of the hedge a costly exercise.

Another approach, pioneered in (Carr & Lee 2004), develops a robust hedging strategy for volatility derivatives which is analogous to the one for variance swaps. In other words the authors find a hedge for a volatility swap using a static position in some European derivative and a dynamic hedging strategy in the underlying. This method works for continuous processes only and is based on the observation that, under the continuity hypothesis, there is a simple algebraic relationship between the Laplace transform of the process and the Laplace transform of its quadratic variation. There are some technical difficulties in computing the relevant integrals for general payoffs of realized volatility. This issue has been dealt with in (Friz & Gatheral 2005) where a formula in terms of Bessel functions is given for the European payoff that one needs to hold in order to hedge the corresponding volatility payoff.

In (Windcliff, Forsyth & Vetzal 2006) the authors investigate hedging techniques for the discretely monitored volatility contracts which are independent of the instantaneous volatility dynamics. Their main result is that the delta hedge of the volatility derivative can be greatly improved by an additional gamma hedge using an at-the-money straddle (or an out-of-the-money strangle) which is re-balanced at each volatility observation time. The reason behind choosing these particular European payoffs lies in the fact that their risk profile resembles that of a volatility swap. Another model-independent hedging approach for variance swaps is presented in (Schoutens 2005). The author shows that, in an environment with jumps, one can use derivatives on the realized higher order moments of the underlying (i.e. the so called moment swaps) to improve the performance of the log contract as a hedge for the variance swap. This strategy provides an improved static hedge, as far as derivatives are concerned, but suffers from the fact that in practice moment swaps of order 3 and above are less liquid than variance swaps themselves.

Another interesting approach to the pricing of volatility derivatives is based on the observation that the term-structure of variance swaps, an example of which is given in figure 19, is mathematically reminiscent of the term-structure of zero-coupon bonds in interest rate modelling. A framework, analogous to the famous HJM, has been proposed in (Buehler 2006). The starting point is the specification of the function-valued process for the forward instantaneous variance which yields an arbitrage-free dynamics for the underlying. This model requires the entire variance swap curve at time zero in order to be calibrated. The correlation between the driving Brownian motion for the stock and the instantaneous variance is used to introduce the ubiquitous skew, but is insufficient to reprice the entire volatility surface that can be observed in the market. Since the driving Markov process for this model is high-dimensional, the pricing is done by Monte Carlo.

This paper is organized as follows. In section 2 we shall describe some of the volatility contracts that can be priced within our framework. Section 3 defines the model for the underlying forward rate. In section 4 we discuss the calibration of the model to a wide range of strikes and maturities for options written on the S&P 500. The key idea, that of the lifting of a Markov generator, which allows us to price general derivatives on the realized variance, is introduced in section 5. The numerical algorithm required to make this idea applicable is described in section 6. Section 7 explains the pricing methodology for derivatives on the realized variance. In section 8 we carry out some numerical experiments and consistency tests on the calibrated model. Concluding remarks are contained in section 9.

2. Volatility derivatives

In this section we are going to give a brief description of the volatility derivatives discussed in this paper. We start with the simplest case, namely a forward on the realized variance, which defines a variance swap. In subsection 2.2 we define options with payoffs that are general functions of realized variance. Subsection 2.3 concerns derivatives that are dependent on implied volatility. In particular we recall the definition of the forward starting options and of the implied volatility index.

2.1. **Variance swaps.** As mentioned above, a variance swap expiring at time T is simply a forward contract on the realized variance Σ_T, quoted in annual terms, of the underlying stock (or index) over the time interval $[0,T]$. The payoff is therefore of the form

$$(\Sigma_T - K_{\text{var}})N,$$

where K_{var} is the strike and N is the notional of the contract. The fair value of the variance is the delivery price K_{var} which makes the swap have zero value at inception.

At present such contracts are liquidly traded for most major indices. The delivery price is usually quoted in the markets as the square of the realized volatility, i.e. $K_{\text{var}} = K^2$ where K is a value of realized volatility expressed in percent. The notional N is usually quoted in dollars per square of the volatility point[3].

A key part of the specification of a variance swap contract is how one measures the realized variance Σ_T. There are a number of ways in which discretely sampled returns of an index (or of an index future[4] F_t) can be calculated and used for defining the realized variance. We will now describe the two most common approaches.

The usual definition of the annualized realized (i.e. accrued) variance of the underlying process F_t in the period $[0,T]$, using logarithmic returns, is $\frac{d}{n}\sum_{i=1}^{n}(\log\frac{F_{t_i}}{F_{t_{i-1}}})^2$, where times t_i, for $i = 0,\dots,n$, are business days from now $t_0 = 0$ until expiry $t_n = T$. The normalization constant d is the number of trading days per year. Another frequently used definition of the realized variance is given by $\frac{d}{n}\sum_{i=1}^{n}(\frac{F_{t_i}-F_{t_{i-1}}}{F_{t_{i-1}}})^2$. It is a standard fact about continuous square-integrable martingales that in the limit, as we make partitions of the interval $[0,T]$ finer and finer, both sums exhibit the following behaviour[5]:

$$\langle \log F\rangle_t = \lim_{n\to\infty}\sum_{i=1}^{n}\left(\log\frac{F_{t_i}}{F_{t_{i-1}}}\right)^2 = \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{F_{t_i}-F_{t_{i-1}}}{F_{t_{i-1}}}\right)^2.$$

The convergence here is in probability and the process $\langle \log F\rangle_t$ is the quadratic-variation[6] process associated to $\log(F_t)$.

In our framework the underlying process F_t will be a continuous-time Markov chain (see chapter 6 of (Grimmett & Stirzaker 2001) for definitions and basic properties). We define the annualized realized variance Σ_T (of F_t over the time interval $[0,T]$) to be the limit

$$\Sigma_T := \frac{1}{T}\lim_{\rho(n)\to 0}\sum_{i=1}^{n}\left(\frac{F_{t_i}-F_{t_{i-1}}}{F_{t_{i-1}}}\right)^2, \tag{1}$$

where, for every $n\in\mathbb{N}$, the set $(t_i)_{i=0,\dots,n}$ is a strictly increasing sequence of times between 0 and T and $\rho(n) := \max\{t_i - t_{i-1};\, i = 1,\dots,n\}$ is the size of the maximal subinterval given by the sequence $(t_i)_{i=0,\dots,n}$ (cf. definitions (18) in section 5 and (19), (20) in subsection 5.1). It should be noted that the techniques described in sections 5 and 6, which provide numerical solubility for our model, can be generalized to the situation where the realized variance is defined as a discretely sampled sum in (1) but without the limit. We are not going to pursue this line of

[3] A *volatility point* is one basis point of volatility, i.e. 0.01 if volatility is quoted in percent. This means that the quote for the notional value of the variance swap tells us how much the swap owner gains if the realized variance increases by $0.0001 = 0.01^2$.

[4] The reason for considering index futures rather than the index itself is twofold. The futures are used for hedging options on the index because they are much easier to trade than the whole portfolio of stocks that the index comprises. Also, it is well-known that futures prices are martingales under the appropriate risk-neutral measure which depends on the frequency of mark-to-market. If the futures contract marks to market continuously, then the price process F_t is a martingale in the risk-neutral measure induced by the money market account as a numeraire. Otherwise we have to take the rollover strategy, with the same frequency as mark-to-market, as our numeraire to obtain the martingale measure for F_t.

[5] Notice also that these equalities hold because the difference of the process $\log(F_t)$ and $\int_0^t \frac{dF_u}{F_u}$ is of finite variation, which is a consequence of Itô's lemma (see theorem 3.3 in (Karatzas & Shreve 1998)).

[6] For a precise definition of a quadratic-variation of continuous square-integrable martingales see chapter 1 of (Karatzas & Shreve 1998).

thought any further, but should notice that the discrete definition of the realized variance would require the application of the block-diagonalization algorithm (section 6) to the probability kernel between any two consecutive observation times t_i rather than the application of the algorithm to the Markov generator directly, which is what is done in section 7.

2.2. **General payoffs of the realized variance.** A *volatility swap* is a derivative given by the payoff

$$(\sigma_T^R - K_{\text{vol}})N,$$

where σ_T^R is the realized volatility over the time interval $[0, T]$ quoted in annual terms, K_{vol} is the annualized volatility strike and N is the notional in dollars per volatility point. The market convention for calculating the annualized realized volatility σ_T^R differs slightly from the usual statistical measure[7] of a standard deviation of any discrete sample and is given by the formula

$$\sigma_T^R = \sqrt{\frac{d}{n}\sum_{i=1}^{n}\left(\frac{F_{t_i} - F_{t_{i-1}}}{F_{t_{i-1}}}\right)^2},$$

where d is the number of trading days per year and t_i are business days from now $t_0 = 0$ until expiry $t_n = T$ of the contract. For our purposes we shall define realized volatility σ_T^R, quoted in annual terms, over the time interval $[0, T]$ as

$$\sigma_T^R := \sqrt{\Sigma_T},$$

where Σ_T is the annualized realized variance defined in (1). It is clear from this definition that the payoff of the volatility swap can be view as a non-linear function of the realized variance.

Since volatility swaps are always entered into at equilibrium, an important issue is the determination of the fair strike K_{vol} for any given maturity T. As discussed in section 1, a term structure of such strikes must be part of the market data that some models require (e.g. (Buehler 2006)) in order to be calibrated. In our case the strikes K_{vol}, for any maturity, are implied by the model which uses as its calibration data the market implied vanilla surface. The value of K_{vol} for a given maturity T is then given by the expectation $\mathbb{E}_0[\sqrt{\Sigma_T}]$, which can easily be obtained as soon as we have the probability distribution function for Σ_T.

The same reasoning applies to variance swaps. The fair strike K_{var} for the variance swap of maturity T can, within our framework, be obtained by taking the expectation $\mathbb{E}_0[\Sigma_T]$. It therefore follows from the concavity of the square root function and Jensen's inequality[8] that the following relationship holds between the fair strikes of the variance and volatility swaps

$$K_{\text{vol}} < \sqrt{K_{\text{var}}},$$

for any maturity T. This inequality is always satisfied by the market quoted prices for variance and volatility swaps and is there to account for the fact that variance is a convex function of volatility. Put differently, this is just a convexity effect, similar to the one observed for ordinary call options, related to the magnitude of volatility of volatility. The larger the "vol of vol" is, the greater the convexity effect becomes. This phenomenon can be observed clearly in the markets with a very steep skew for implied volatilities. If one wanted to estimate its size in general, it would be necessary to make assumptions about both the level and volatility of the future realized volatility. Within our model this can be achieved directly by comparing the values of the two expectations (see figure (19) for this comparison based on the market implied vanilla surface for the S&P 500).

[7]Given a sample of n values $X_1, \ldots, X_n$ with the mean $\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$, the unbiased statistical estimation of the standard deviation is given by

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2}.$$

[8]For any convex function $\phi : \mathbb{R} \to \mathbb{R}$ and any random variable $X : \Omega \to \mathbb{R}$ with a finite first moment Jensen's inequality states that $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$.

There are other variance payoffs which are of practical interest and can be priced and hedged within our framework. Examples are volatility and variance swaptions whose payoffs are $(\sqrt{\Sigma_T} - K_{\text{vol}})^+$ and $(\Sigma_T - K_{\text{var}})^+$ respectively, where as usual $(x)^+$ equals $\max(x, 0)$ for any $x \in \mathbb{R}$. Capped volatility swaps are also traded in the markets. Their payoff function is of the form $(\min(\sqrt{\Sigma_T}, \sigma_m) - K_{\text{vol}})$, where σ_m denotes the maximum allowed realized variance. It is clear that all such contracts can be priced easily within our framework by integrating any of these payoffs against the probability distribution function (see figure 11 for maturities above 6 months) of the annualized realized variance Σ_T and then multiplying the expectation with the corresponding discount factor.

It should be noted that there exist even more exotic products, like corridor variance swaps (see (Carr & Lewis February 2004)), whose payoffs depend on the variance that accrues only if the underlying is in a predefined range. Such products cannot be priced directly in the existing framework. A minor modification of the model would be required to deal with this class of derivatives. However we will not pursue this avenue any further.

2.3. **Forward staring options and the volatility index.** Let T' and T be a pair of maturities such that $T' < T$. A *forward staring option* (or a *forward-start*) is a vanilla option with expiry T and the strike,set at time T', which is equal to $\alpha S_{T'}$. The quantity S_t is the underlying financial instrument the option is written on (usually a stock or an index). More formally the value of a forward-start at time T (i.e. its payoff) is given by

$$V_{FS}(T) = (S_T - \alpha S_{T'})^+, \tag{2}$$

where the constant α is specified at the inception of the contract and is know as the *forward strike*. It is clear from the definition of the forward-start that its value at time T' equals the value of a plain vanilla call option

$$V_{FS}(T') = V_C(S_{T'}, T - T', \alpha S_{T'})$$

that expires in $T - T'$ years and whose strike equals $\alpha S_{T'}$. Notice that at time T' the constant α can be characterized as the ratio[9] between the spot price $S_{T'}$ and the strike of the call option into which the forward-start is transformed. In the classical Black-Scholes framework, we have an explicit formula, denoted by $\text{BS}(S_{T'}, T - T', \alpha S_{T'}, r', \sigma')$, for the value of this call option (see (Black & Scholes 1973)). This formula depends linearly on the spot level $S_{T'}$ if the ratio of the spot and the strike (i.e. the "moneyness") is known. In other words, assuming we are in the Black-Scholes world with a deterministic term-structure of volatility and interest rates and zero dividends, we can express the value of the forward-start at time T' as

$$V_{FS}(T') = S_{T'}\text{BS}(1, T - T', \alpha, r', \sigma'),$$

where σ' is the forward volatility rate[10] and r' is the forward interest rate between T' and T. The following key observations about the Black-Scholes value and sensitivities of the forward starting option are now clear:

- the value equals $V_{FS}(0) = S_0\text{BS}(1, T - T', \alpha, r', \sigma')$ and the delta (i.e. $\frac{\partial}{\partial S_0} V_{FS}(0)$) is simply $\text{BS}(1, T - T', \alpha, r', \sigma')$,
- the forward starting option is gamma neutral[11] (i.e. $\frac{\partial^2}{\partial S_0^2} V_{FS}(0) = 0$) and
- the contract has non-zero vega (i.e. $\frac{\partial}{\partial \sigma'} V_{FS}(0) > 0$).

We are interested in the Black-Scholes pricing formula for the forward-starts because we need to use it when expressing the forward volatility smile of our model. The values of forward volatility σ', implied by the equation $V_{FS}(0) = S_0\text{BS}(1, T - T', \alpha, r', \sigma')$, are plotted in figures 7,

[9] This ratio is sometimes referred to as the "moneyness" of the option.

[10] Assuming that the term-structure of volatility is parametrized by $\sigma(t)$, the forward volatility rate σ' is given by $\sigma'^2 = \frac{1}{T - T'} \int_{T'}^{T} \sigma(t)^2 dt$.

[11] It should be noted that, in the presence of stochastic volatility, the gamma of a forward starting option is no longer necessarily zero. However in a realistic model it should not be too large because it reflects the dependence of volatility on very small moves of the underlying, the effect of which should be negligible.

8, 9 and 10 for a wide range of values of the forward strike α and a variety of time horizons T', T. In other words, we first calculate the value $V_{FS}(0)$ of the forward-start and then invert the Black-Scholes pricing formula to obtain the implied forward volatility σ'.

We shall now give a brief description of the implied volatility index (VIX) and then move on to discuss the future probability distribution of VIX, which, as will be seen, is directly related to forward starting options.

VIX was originally introduced in 1993 by Chicago Board Options Exchange (CBOE) as an index reflecting the 1 month implied volatility of the at-the-money put and call options on S&P 100. To facilitate trading in VIX, in 2003 CBOE introduced a new calculation, using a full range of strikes for out-of-the-money options, to define the value of VIX. At the same time the underlying financial instrument on which the options are written was changed to S&P 500 (for a detailed description of these changes and their ramifications see (CBOE 2003)). The new formula is

$$\sigma^2_{\text{VIX}} = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T}\left(\frac{F}{K_0} - 1\right)^2, \tag{3}$$

where the index itself is given by VIX $= 100\sigma_{\text{VIX}}$. The sequence K_i consists of all the exchange quoted strikes and the quantities $Q(K_i)$ are the corresponding values of out-of-the-money put/call options expiring at maturity T, where T equals 1 month[12]. The quantity F in formula (3) is the forward value of the S&P 500 index derived from option prices[13] and the at-the-money strike K_0 is defined as the largest strike below F. Note also that it is precisely at K_0 that the symbol $Q(K)$ in (3) changes from put to call options.

The reason why formula (3) allows easier trading of volatility follows from the simple observation that σ^2_{VIX} is essentially the value of a European derivative, expiring at time T, with the logarithmic payoff given in (4). This is a consequence of the well-known decomposition of any twice differentiable payoff described in (Breeden & Litzenberger 1978), (Carr & Madan 1998), (Demeterfi et al. 1999*b*) and other sources:

$$-\log\left(\frac{S_T}{F}\right) = -\frac{S_T - F}{F} + \int_0^F \frac{1}{K^2}(K - S_T)^+ dK + \int_F^\infty \frac{1}{K^2}(S_T - K)^+ dK. \tag{4}$$

This formula holds for any value of F, but expressions simplify if we assume that F equals the forward of the index S_t at time T (i.e. $F = \mathbb{E}[S_T]$). By taking the expectation with respect to the risk-neutral measure we get the following expression for the forward price of the log payoff:

$$-\mathbb{E}\left[\log\left(\frac{S_T}{F}\right)\right] = \int_0^F \frac{1}{K^2} e^{rT} P(K) dK + \int_F^\infty \frac{1}{K^2} e^{rT} C(K) dK, \tag{5}$$

where $C(K) = e^{-rT}\mathbb{E}[(S_T - K)^+]$ (resp. $P(K) = e^{-rT}\mathbb{E}[(K - S_T)^+]$) is the price of a call (resp. put) option struck at K. It is shown in (Demeterfi et al. 1999*b*) that the portfolio of vanilla options given by (5) can be used to hedge perfectly a variance swap if there are no jumps in the underlying market. From our point of view the expression (5) is interesting because a simple calculation shows that definition (3) is a possible discretization of it. By defining its own version of the approximation to the logarithm, CBOE has created a volatility index which can be replicated by trading a relatively simple European payoff. This feature greatly simplifies the trading of VIX.

Since the implied volatility index is defined by a portfolio of puts and calls in (3), it is clear that the random nature of the value of VIX at time t will be determined by the value of the

[12]The actual CBOE definition of VIX uses two maturities, rather than one, and the two corresponding strips of options. The value of σ^2_{VIX} is then defined to be a convex combination of the two values given by the formula in (3). Details of this construction can be found in (CBOE 2003). We are going to neglect this technical point because its use is mainly to circumvent certain market irregularities when dealing with options with short time to expiry and does not add complexity to the modelling side of the problem.

[13]See page 3 in (CBOE 2003) for the precise definition.

corresponding portfolio of forward starting options. The probability distribution function for the behaviour of the volatility index at t is obtained from the model by the following procedure:

(I) Fix a level S of the underlying.
(II) Find the probability that the price process is at level S at a given time t in the future.
(III) Evaluate the portfolio of options that define the volatility index between times t and $t+T$, conditional on the process S_t being at level S.
(IV) Repeat these steps for all attainable levels S for the underlying Markov chain S_t.
(V) Subdivide the real line into intervals with disjoint interiors of length δ, where δ is a small positive number. To each of the intervals assign a probability that is a sum of probabilities in step (II) corresponding to the values obtained in step (III) that lie within the interval.

This describes the construction of the probability distribution function of the volatility index at time t. The plot of this pdf for a variety of maturities can be seen in figure 20.

Our final task is to price any European payoff written on the level of VIX at a certain time horizon. Given that our model allows us to extract the pdf of VIX for any expiry, pricing such a derivative amounts to integrating the payoff function against the probability distribution that was described above.

3. The model for the underlying

In this section we are going to describe the model for the underlying equity index which we later calibrate to the implied volatility surface for vanilla options on the S&P 500 (see section 4 and figure 1). Our model will be a mixture of local and stochastic volatility coupled with an infinite activity jump process and will be defined on a continuous-time lattice in a largely non-parametric fashion. The basic framework is an application to equity derivatives of the methods used in (Albanese & Mijatović 2005) for modelling the foreign exchange rate. Our basic tools for all the constructions that follow will be spectral theory and numerical linear algebra.

We are assuming that, apart from the options data, we are also given a term-structure of interest rates and a continuously compounded deterministic dividend schedule. In other words interest rate $r(t)$ and dividend yield $d(t)$ are given as deterministic functions of calendar time t. Our modelling primitive will not be the equity index S_t itself but its forward price $F_t = e^{(d(t)-r(t))t}S_t$. Put differently, the model will be defined under the forward measure with the underlying process a martingale, because, on a lattice, it is numerically more convenient to simulate a stochastic process without drift (i.e. a martingale) than one with a drift.

The process for the forward price F_t will be defined as follows. We shall first introduce a number of local volatility regimes, all following CEV processes with different parameter sets. We will then add jumps to each of them using subordination and then introduce the stochasticity of volatility by specifying its dynamics so that it is correlated to the level of the forward. In the rest of this section we will go through these steps in more detail. In subsection 3.5 we will address the issue of the pricing and hedging of European derivatives and forward starting options.

3.1. The local volatility processes.

Our model will comprise M local volatility regimes (in order to fit the implied volatility surface for the S&P 500, we used $M = 5$ regimes). The switching between the volatility regimes will be driven by a stochastic process, which will be correlated with the level of the forward price F_t. This construction will be carried out in subsection 3.3.

As mentioned above, all these local volatility regimes will be from the same class, namely that of CEV processes. We are now going to define a continuous-time Markov chain, which is a discretization of a generic CEV process. The family of local volatility processes will then be obtained by judiciously choosing a family of the CEV parameters.

Recall that forward price F_t can be defined as a CEV process by the following stochastic differential equation

$$dF_t = v(F_t)dW_t, \quad \text{where} \quad v(F_t) := \min(\sigma F_t^{\beta}, \bar{\sigma}) \tag{6}$$

and where W_t is the standard Brownian motion. The capping constant $\bar{\sigma}$ was introduced into this definition of the CEV process in order to avoid accumulation of probability at the boundary of the domain of the process F_t. It is well-known (e.g. see chapter 5 in (Karatzas & Shreve 1998)) that the Markov generator of a diffusion process given by the stochastic differential equation (6) acts in the following way on any twice differentiable real function ϕ:

$$(\mathcal{L}\phi)(F) = \frac{v(F)^2}{2}\frac{\partial^2\phi}{\partial F}(F). \tag{7}$$

The probability density function $p(G,t;F,T) = \mathbb{P}(F_T = F|F_t = G)$ is the solution of the partial differential equation

$$\frac{\partial p}{\partial t} + \mathcal{L}p = 0$$

with boundary condition $p(G,T;F,T) = \delta(G-F)$, where δ is the Dirac delta function. This PDE is known as the *backward Kolmogorov equation* for the process F_t and it implies that all the information required to obtain the probability kernel $p(G,t;F,T)$ is contained in the Markov generator given by (7). For this reason we will define the probability kernel of the continuous-time Markov chain that will approximate the process F_t using a natural discretization of the Markov generator $\mathcal{L}$.

In order to define this dynamics we must first define a discrete domain for the forward rate F_t which can be achieved as follows. Let Ω be a finite set $\{0,\ldots,N\}$ containing the first N integers together with 0 and let $F : \Omega \to \mathbb{R}$ be a non-negative function which satisfies the following two conditions: $F(0) \geq 0$ and $F(x) > F(x-1)$ for all x in $\Omega - \{0\}$. Given such a function F, the discretized forward rate process F_t^Ω can take any of the values $F(x)$, where x is an element in Ω and time t is smaller than some time horizon T.

The next step is to ensure that the dynamics of the discretized forward process correspond to the dynamics specified by the stochastic differential equation (6). As mentioned above this will now be achieved by reinterpreting the Markov generator (7) in the discrete setting. Note first that the differential operator $\mathcal{L}$ is just the Laplace operator multiplied by a scalar function. A natural discretization of the Laplace operator is given by

$$(\Delta^\Omega u)(x) := \begin{cases} u(x+1) + u(x-1) - 2u(x) & \text{if } x \in \Omega - \{0,N\}, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

for all functions u on Ω. Note that this definition of the discretized Laplace operator imposes absorbing boundary conditions on the Markov chain, associated to Δ^Ω, at each end of the domain Ω. This is a reasonable requirement of the underlying process for two reasons. Firstly, since the size of the set Ω is a parameter of our model, we can make sure that it is of sufficient size so that we can calibrate to the market data without the process ever reaching the boundary. Secondly, this choice of boundary conditions makes it easy to detect if the domain Ω is not large enough, which would not necessarily be the case had we used reflecting boundary conditions[14].

Let the operator $\mathcal{L}^\Omega$ denote the discretized version of the generator $\mathcal{L} = \frac{v^2}{2}\Delta$. Using the discrete version of the Laplace operator (8), we can define the generator $\mathcal{L}^\Omega$ as a tridiagonal matrix of size $(N+1)\times(N+1)$ with entries $\mathcal{L}^\Omega(x,y)$, where x and y are elements of Ω, that satisfy the following conditions for all x in Ω:

$$\sum_{y\in\Omega} \mathcal{L}^\Omega(x,y) = 0, \tag{9}$$

$$\sum_{y\in\Omega} \mathcal{L}^\Omega(x,y)(F(y)-F(x)) = 0, \tag{10}$$

$$\sum_{y\in\Omega} \mathcal{L}^\Omega(x,y)(F(y)-F(x))^2 = v(F(x))^2. \tag{11}$$

[14]The choice of the boundary behaviour of a Markov chain representing the realized variance of the underlying will be of fundamental importance in sections 5 and 6. There however, a different, perhaps slightly unnatural, boundary condition will be most useful.

Condition (9) is there to secure probability conservation over the infinitesimal time interval dt and stems from the fact that a generator of a continuous-time Markov chain consists of the gradients of probability for jumping from any point $x \in \Omega$ to any other point $y \in \Omega$. Therefore equation (9) is simply a derivative of the equation $\sum_{y\in\Omega} \mathbb{P}(F_s^\Omega = F(y)|F_t^\Omega = F(x)) = 1$ with respect to s at time t. The second and the third conditions are the instantaneous[15] first and second moment-matching conditions for the discretized process F_t^Ω. Equality (10) is a martingale condition for F_t^Ω since the process F_t has no instantaneous drift. Condition (11) guarantees that the Markov chain has the same instantaneous variance as the diffusion defined by the SDE in (6).

The final constraint in the construction of the Markov generator $\mathcal{L}^\Omega$ comes from the specification of the process at the boundary of its domain. We would like to ensure that the process F_t^Ω obeys absorbing boundary conditions whenever it gets that far. As discussed above, this can be done by setting all elements in the top and bottom row of the matrix $\mathcal{L}^\Omega$ to be equal to zero. In coordinates this can be expressed by the condition

$$\mathcal{L}^\Omega(x, y) = 0, \quad \text{for all} \quad y \in \Omega, \ \ x \in \{0, N\}.$$

Notice that this operation does not interfere with conditions (9), (10) and (11) and hence gives us a well-defined generator for the Markov process F_t^Ω.

3.1.1. *Computing the probability kernel.* Our next task is to obtain the probability kernel of the process F_t^Ω from its Markov generator $\mathcal{L}^\Omega$. This can be achieved for very general Markov processes by applying spectral methods of operator theory. Here we will only illustrate the spectral resolution method in the special case of the finite-dimensional matrix $\mathcal{L}^\Omega$. This method is sufficient for our purposes and can be applied directly to other cases which are of interest to us, such as the introduction of jumps (see subsection 3.2) and the calculation of the probability kernel for the realized variance of the underlying process (see section 7). We start by considering the following eigenvalue problem

$$\mathcal{L}^\Omega u_n = \lambda_n u_n \tag{12}$$

for the matrix $(\mathcal{L}^\Omega(x, y))_{x,y\in\Omega}$. The vectors u_n are the eigenvectors of the linear operator $\mathcal{L}^\Omega$ and the scalars λ_n are the corresponding eigenvalues. Except in trivial cases, the Markov generator $\mathcal{L}^\Omega$ will not be a symmetric matrix, which implies that the zeros of the characteristic polynomial of $\mathcal{L}^\Omega$ (i.e. the eigenvalues λ_n) will not be real. On the other hand it is not hard to see that an element λ_n of the spectrum of any Markov generator must have a non-positive real part ($\mathcal{R}e(\lambda_n) \leq 0$) and that the complex eigenvalues occur in conjugate pairs (i.e. λ_n is an eigenvalue if and only if $\overline{\lambda_n}$ is an eigenvalue).

In general of course, there is no guarantee that there exists a complete set of $(N + 1)$ eigenvectors u_n for the operator $\mathcal{L}^\Omega$. However, such a set will certainly exist if we can find $(N + 1)$ distinct eigenvalues λ_n of $\mathcal{L}^\Omega$. But the set of all $(N + 1) \times (N + 1)$ matrices which do not have distinct eigenvalues must have Lebesgue measure zero for the same reason that the set of all polynomials of order $(N + 1)$ with at least two coinciding zeros has Lebesgue measure zero. Therefore the chance that such a complete set does not exist, for a Markov generator specified in a nonparametric fashion is zero, so we can safely assume that the complete set of eigenvectors exists. In the unlikely event that this assumption is not valid, the numerical linear algebra routines needed to solve our lattice model will identify the problem and an arbitrarily small perturbation of a given operator will suffice to rectify the situation. Assuming that there is a solution, the diagonalization problem (12) can be rewritten in the following matrix form

$$\mathcal{L}^\Omega = U\Lambda U^{-1},$$

where U is the matrix whose columns are the eigenvectors u_n and Λ is the diagonal matrix with the eigenvalues λ_n.

[15]It is well-known that the first two instantaneous moments of any diffusion determine its finite-dimensional distribution functions completely (see (Karatzas & Shreve 1998)). It is therefore sufficient for the process F_t^Ω to match these two conditions in order to be a valid approximation of F_t.

Key to our constructions is the remark that, if the Markov generator is diagonalizable, we can apply to it an arbitrary function ϕ, defined on the spectrum of the generator, by means of the following formula:

$$\phi(\mathcal{L}^\Omega) = U\phi(\Lambda)U^{-1}. \tag{13}$$

The expression (13) is useful because the task of calculating $\phi(\Lambda)$ is a very simple one:

$$\phi(\Lambda) = \begin{pmatrix} \phi(\lambda_0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi(\lambda_N) \end{pmatrix}.$$

It can be seen by a direct calculation (see chapter 6 in (Grimmett & Stirzaker 2001)) that the matrix $P_t = e^{(T-t)\mathcal{L}^\Omega}$ satisfies the backward Kolmogorov equation, much like the one for the original diffusion,

$$\frac{\partial P_t}{\partial t} + \mathcal{L}^\Omega P_t = 0 \tag{14}$$

with the boundary condition $P_T = \mathbb{I}$, where $\mathbb{I}$ denotes the identity matrix on the vector space $\mathbb{R}^N$. This fact, combined with formula (13), gives us an explicit expression for the transition probability kernel of F_t^Ω:

$$\mathbb{P}(F_T^\Omega = F(y)|F_t^\Omega = F(x)) = (e^{(T-t)\mathcal{L}^\Omega})(x,y) = \sum_{n=0}^{N} e^{\lambda_n(T-t)} u_n(x)v_n(y).$$

In this expression the vectors v_n correspond to the columns of the matrix U^{-1}.

3.2. **Adding jumps.** As defined so far the Markov chain F_t^Ω behaves like a pure diffusion process. It is known that this class of models is not well suited to explain the volatility skew for options with short maturity because of the extremely small probabilities of large moves in the underlying in short-time horizons. In the equity markets however, jumps are commonplace and as such they influence the prices of short-dated out-of-the-money options. If we want to calibrate our model to the entire volatility surface we therefore need to introduce jumps into the risk-neutral dynamics of the underlying forward rate F_t^Ω. Using spectral theory this can be easily achieved in a general way. What we want is to have different distributions of jump sizes for jumps up and jumps down. Having this property in our model is crucial because the market expectations for jumps up and down are know to the market makers and are almost always very different from each other. Therefore, any process that aspires to model the risk-neutral dynamics of the underlying correctly must be able to account for this difference. The variance-gamma model, defined in (Madan, Carr & Chang 1998), has this property since the characteristic function of the underlying process[16] is not real.

A general way of building infinite-activity jump processes is by subordinating diffusions using a special class of stochastic time changes. Such a time change is given by a non-decreasing stationary process T_t with independent increments. The time change T_t is known as a *Bochner subordinator* and is characterized by a *Bernstein function* $\phi(\lambda)$ which has the following property

$$E_0\left[e^{-\lambda T_t}\right] = e^{-\phi(\lambda)t}.$$

In other words the process T_t is a non-decreasing stationary process with independent increments whose Laplace transform takes the special form $e^{-\phi(\lambda)t}$, where $\phi(\lambda)$ is the Bernstein function of the process.

For example in the case of the afore mentioned variance-gamma process, the Bernstein function is of the form

$$\phi(\lambda) = \frac{\mu^2}{\nu}\log\left(1 + \lambda\frac{\nu}{\mu}\right). \tag{15}$$

[16]The process used in (Madan et al. 1998) is a time-changed Brownian motion with drift. The stochastic time is given by a gamma process.

The parameter μ is the mean-reversion rate (usually taken to be equal to 1) and ν is the variance rate of the variance-gamma process.

It was shown in (Phillips 1952) that, given a general Markov process X_t with a generator $\mathcal{L}$ and a Bochner subordinator T_t, the subordinated process X_{T_t} is a Markov process with a generator $\mathcal{L}' = -\phi(-\mathcal{L})$, where ϕ is the Bernstein function of T_t.

In our framework we need to add jumps to the local volatility process F_t^Ω given by the generator $\mathcal{L}^\Omega$. In order to produce asymmetric jumps we specify two Bernstein functions of the variance-gamma process by choosing two jump intensities ν_+ for up and ν_- for down.

We then compute separately the two Markov generators

$$\mathcal{L}_\pm = -\phi_\pm(-\mathcal{L}^\Omega) = -U_\pm \phi_\pm(-\Lambda) U_\pm^{-1},$$

where Λ is the diagonal matrix from subsection 3.1.1 and the Bernstein functions $\phi_\pm$ are given by

$$\phi_\pm(\lambda) = \frac{1}{\nu_\pm} \log(1 + \lambda \nu_\pm).$$

Each of the square matrices $\mathcal{L}_\pm$ corresponds to a time-changed diffusion process. In particular the elements of $\mathcal{L}_+$, which are above the diagonal, are the (scaled) probabilities of jumping up in the infinitesimal time interval dt. On the other hand, the sub-diagonal triangle of $\mathcal{L}_-$ contains the (scaled) probabilities of jumping down in the time interval dt. Thus we can define a new generator for our process, which will have asymmetric jumps, by combining the two generators in the following way

$$\mathcal{L}^\Omega = \begin{pmatrix} d(0,0) & \mathcal{L}_+(0,1) & \cdots & \mathcal{L}_+(0,N-1) & \mathcal{L}_+(0,N) \\ \mathcal{L}_-(1,0) & d(1,1) & \cdots & \mathcal{L}_+(1,N-1) & \mathcal{L}_+(1,N) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{L}_-(N-1,0) & \mathcal{L}_-(N-1,1) & \cdots & d(N-1,N-1) & \mathcal{L}_+(N-1,N) \\ \mathcal{L}_-(N,0) & \mathcal{L}_-(N,1) & \cdots & \mathcal{L}_-(N,N-1) & d(N,N) \end{pmatrix}.$$

Since we want our new process with jumps to be a martingale, we need to make sure that condition (10) is satisfied for the new generator $\mathcal{L}^\Omega$. This can be done easily by adjusting the elements just above and below the diagonal of the matrix $\mathcal{L}^\Omega$. If for example the drift in the x-th row of $\mathcal{L}^\Omega$ is positive, we add some probability to the element $\mathcal{L}^\Omega(x, x-1)$ so that the condition (10) becomes valid. If, on the other hand, the drift in the x-th row is negative, then we can help the generator pull up the process by adding probability to the element $\mathcal{L}^\Omega(x, x+1)$. Once we do this for all x, the new modified operator, which we again call $\mathcal{L}^\Omega$, will satisfy the martingale condition in (10).

A possible undesirable effect of subordination on the underlying process is a linear deterministic time change (along with the stochastic one). This would result in the distortion of the instantaneous variance of the underlying process by a constant factor. In order to avoid this effect, we multiply each row of the subordinated generator by a constant chosen in such a way, so that equation (11) holds. Since the function $v(F(x))$ on the righthand side of (11) gives the value of the instantaneous variance of the local volatility process we started with, this procedure eliminates the linear deterministic time change that might have occurred during subordination. Notice also that the rescaling of the rows of $\mathcal{L}^\Omega$ has preserved the martingale condition in (10), which has been manufactured above.

The procedures we have just carried out did not require the knowledge of the diagonal elements of $\mathcal{L}^\Omega$. They need to be chosen in such a way that the probability conservation (condition (9)) is satisfied. This can be achieved by simply defining the diagonal elements in the following way

$$d(x,x) = -\sum_{y \in \Omega - \{x\}} \mathcal{L}^\Omega(x,y).$$

This gives us a well-defined Markov generator $\mathcal{L}^\Omega$ for a diffusion process with jumps, which can be used to model the risk-neutral forward rate because it is a martingale, and whose instantaneous variance has not been altered by subordination.

The construction we have just presented differs in two ways from the one in (Madan et al. 1998). Firstly our time-changing approach is applicable to a general diffusion process which need not be translation invariant[17]. Secondly we have the additional flexibility of specifying explicitly the intensities of up-jumps and down-jumps separately.

3.3. **Modelling the dynamics of stochastic volatility.** In order to introduce the stochasticity of volatility into our modelling framework, we need to start with $M \in \mathbb{N}$ jump-diffusions of subsection 3.2 (all defined on the same domain Ω) which are given by Markov generators $\mathcal{L}^{\Omega}_{\alpha}$ for $\alpha = 0, \ldots, M-1$. In this subsection we are going to define the dynamics of stochastic volatility that will give our model the ability to switch between the various jump-diffusion regimes. As will become clear, the stochasticity of volatility will be sensitive to the current level of the underlying, thus making it possible to relate the model to a particular view of the market. It is not uncommon among market participants to express market views (i.e. skew and smile behaviour of option prices) in terms of the future "levels" at which the underlying might be trading. In our framework we can set these levels explicitly and are allowed to choose independently the corresponding jump-diffusion, that can express the required view of the volatility surface. The introduction of stochastic volatility into the model will be in several stages. Let us start by specifying the dynamics of stochastic volatility and then combining them with the underlying jump-diffusion regimes to give a full specification of the model.

Let V be the set $\{0, \ldots, M-1\}$ of all possible volatility states. For each volatility state γ (in V) we define a Markov generator $\mathcal{G}^{V}_{\gamma}$ by specifying the matrix elements $\mathcal{G}^{V}_{\gamma}(\alpha, \beta)$, for all $\alpha, \beta \in V$, so that the continuous-time diffusion given by $\mathcal{G}^{V}_{\gamma}$ mean-reverts to one of the volatility states in V.

We have defined M Markov generators $\mathcal{G}^{V}_{\gamma}$, each of them specifying its own dynamics of the stochastic volatility process. Our next task is to obtain a single global stochastic volatility Markov generator which will favour a certain regime γ conditional on the position of the forward rate. This can be achieved by using a *partition of unity* which is described as follows. Choose a strictly increasing sequence F_{γ} of the forward rate levels so that, if the forward process is close to the level F_{γ}, the market views of the smile and skew agree with the ones implied by the process $\mathcal{L}^{\Omega}_{\gamma}$ from subsection 3.2. The partition of unity is defined as a sequence of M functions $\epsilon_{\gamma} : \mathbb{R} \to [0, 1]$ with the key property

$$\sum_{\gamma=0}^{m-1} \epsilon_{\gamma}(F) = 1, \quad \text{for all} \quad F \in \mathbb{R}.$$

Given the sequence of levels F_{γ}, such functions can be defined explicitly, using a piecewise linear scheme, in the following way:

$$\epsilon_{\gamma}(F) := \begin{cases} \frac{F - F_{\gamma-1}}{F_{\gamma} - F_{\gamma-1}} & F \in [F_{\gamma-1}, F_{\gamma}] \\ \frac{F_{\gamma+1} - F}{F_{\gamma+1} - F_{\gamma}} & F \in [F_{\gamma}, F_{\gamma+1}] \\ 0 & \text{otherwise.} \end{cases}$$

This definition has to be modified slightly for the boundary cases when γ equals 0 or $(M-1)$:

$$\epsilon_0(F) := \begin{cases} 1 & F \leq F_0 \\ \frac{F_1 - F}{F_1 - F_0} & F \in [F_0, F_1] \\ 0 & F \geq F_1, \end{cases} \qquad \epsilon_{m-1}(F) := \begin{cases} 0 & F \leq F_{m-2} \\ \frac{F - F_{m-2}}{F_{m-1} - F_{m-2}} & F \in [F_{m-2}, F_{m-1}] \\ 1 & F \geq F_{m-1}. \end{cases}$$

We are now in a position to define our global Markov generator for the stochastic volatility process, which has the capability of changing its properties when the forward rate undergoes a

[17] A key feature of Lévy processes is that they are translation invariant. That is precisely the property of the Brownian motion with a drift required in the construction of the jump process in (Madan et al. 1998).

substantial move. The definition, using the partition of unity, goes as follows

$$\mathcal{L}_x^V(\alpha,\beta) := \sum_{\gamma=0}^{m-1} \epsilon_\gamma(F(x))\mathcal{G}_\gamma^V(\alpha,\beta),$$

where α, β are elements in V and $F(x)$ is the function on Ω describing the forward rate process. It follows from the defining property of partition of unity that the matrix $\mathcal{L}_x^V$ is indeed a Markov generator for any element x of the underlying space Ω.

Just like with any stochastic volatility model, our aim is to define a Markov generator $\mathcal{L}$ which will specify the probabilities of going from any state (x,α) in $\Omega \times V$ to any other state (y,β) (of the same set) in the infinitesimal time interval dt. The Markov generator $\mathcal{L}$ that specifies the dynamics of our model for the underlying forward rate is given by

$$\mathcal{L}(x,\alpha;y,\beta) := \mathcal{L}_\alpha^\Omega(x,y)\delta_{\alpha\beta} + \mathcal{L}_x^V(\alpha,\beta)\delta_{xy}, \tag{16}$$

where δ denotes the Kronecker delta function. Note that it follows trivially, from the properties of the Kronecker delta, that the matrix $\mathcal{L}$ is a genuine Markov generator (i.e. it has positive entries off the diagonal and its rows sum to one). Another important feature is that the generator $\mathcal{L}$, by definition, does not allow for simultaneous jumps of the state and the volatility variables. This property ensures that our forward process F_t^Ω, whose dynamics is specified by $\mathcal{L}$, remains a martingale:

$$\begin{aligned}\mathbb{E}_t^{(x,\alpha)}[dF_t^\Omega] &= \sum_{(y,\beta)\in\Omega\times V} (F(y)-F(x))\left(\mathcal{L}_\alpha^\Omega(x,y)\delta_{\alpha\beta} + \mathcal{L}_x^V(\alpha,\beta)\delta_{xy}\right)\\ &= \sum_{y\in\Omega}(F(y)-F(x))\,\mathcal{L}_\alpha^\Omega(x,y) + (F(x)-F(x))\sum_{\beta\in V}\mathcal{L}_x^V(\alpha,\beta) = 0.\end{aligned}$$

3.4. **Deterministic time-change.** The model we have described so far is completely stationary. This amounts to the fact that the implied volatility surface is influenced purely by the model parameters and has no explicit time dependence. By adjusting these parameters one can obtain a general shape of the market impled volatility surface and reprice correctly the out-of-the-money options.

In order to get a good match for the term structure of the at-the-money implied volatilities we have to introduce a minimal deterministic time-change, which only has a marginal effect on the out-of-the-money options. This adjustment can be specified by an increasing function $f : [0,T] \to [0,\infty)$ which deterministically transforms calender time t to financial time $f(t)$. Since our model can capture well the features of the underlying market before the deterministic time-change is introduced, the calender time t and the financial time $f(t)$ will differ only slightly from each other. We first find values of the function f for all market specified maturities by satisfying the requirement that the at-the-money options are priced correctly. It will become clear in subsection 3.5 that a deterministic time-change plays a very isolated role in the pricing algorithm. Therefore finding a correct value of f at the market specified maturities is not a difficult task. For all other times t we use a linear interpolation to arrive at the value $f(t)$. The graph of the function f, used to calibrate the model to the market implied volatility surface for S&P 500, is given in figure 3.

3.5. **Pricing and hedging of vanilla options and forward-starts.** The pricing problem for a general European option expiring at time T reduces to the calculation of the transitional probability density function $p((x,\alpha),t;(y,\beta),T)$ for the underlying process F_t^Ω, which was defined in subsection 3.3 via its Markov generator (16) (recall that $(x,\alpha),(y,\beta)$ are elements of $\Omega \times V$ while t is the current time). This is because our framework is defined in the risk-neutral measure which implies that the value of any security at time t equals the discounted expectation of the value of the same security at any time horizon T.

Since we would like to include the deterministic time-change of subsection 3.4 in the definition of the density $p((x,\alpha),t;(y,\beta),T)$, we must start by specifying the risk-neutral dynamics of the

underlying forward process in financial time. More precisely let

$$U_s = (u((x,\alpha), s; (y,\beta), S))_{(x,\alpha),(y,\beta)\in\Omega\times V}$$

be a stochastic matrix induced by the Markov generator $\mathcal{L}$. In other words the coordinate functions $u((x,\alpha), s; (y,\beta), S)$ are the transition probabilities to go between the states (x,α) and (y,β) in the time interval $[s,S]$. The quantity s denotes the financial time and can therefore be expressed as $s = f(t)$, where f is the function from subsection 3.4 and t is the calendar time. Similarly $S = f(T)$ is the financial time horizon.

As was mentioned in subsection 3.1, the matrix U_s satisfies the backward Kolmogorov[18] equation (14) with the boundary condition $U_S = \mathbb{I}$ ($\mathbb{I}$ is the identity matrix on the vector space $\mathbb{R}^k$ where $k = m(N+1)$). The solution of the backward Kolmogorov equation can therefore be expressed in terms of functional calculus as $U_s = \exp((S-s)\mathcal{L})$ and can be calculated explicitly using the spectral decomposition of the operator $\mathcal{L}$ (which in this case is just $k \times k$ matrix) as given by (13). The first step is to calculate the eigenvalues λ_n of the Markov generator $\mathcal{L}$ and the second is to find the eigenvectors u_n of $\mathcal{L}$, put them into a transition matrix U, and find the columns of the inverse U^{-1} which we will denote by v_n. In section 4 we calibrate the model using a lattice with $M(N+1) = 5 \cdot 76 = 380$ points which implies that the generator $\mathcal{L}$ is a square matrix of dimension 380×380. For matrices of this size diagonalization routines such as `dgeev` in the numerical linear algebra library LAPACK are very efficient.

Once we have the spectral decomposition of $\mathcal{L}$, we can calculate the probability kernel $p((x,\alpha), t; (y,\beta), T)$, which depends on calender time, using functional calculus:

$$p((x,\alpha), t; (y,\beta), T) = e^{(f(T)-f(t))\mathcal{L}}((x,\alpha),(y,\beta)) = \sum_{n=1}^{M(N+1)} e^{\lambda_n(f(T)-f(t))} u_n(x,\alpha) v_n(y,\beta).$$

Since we know that eigenvalues λ_n must have a negative real part, in the case of long-dated options only very few eigenvalues will play a role, because the exponential of a negative number becomes negligibly small very quickly. Another surprising and important fact which follows from the above representation of the probability kernel is that it depends in a very isolated way on the (financial) time to maturity.

The current price C_t of a European derivative paying $h(S_T)$ at the time horizon T, where S_t is the underlying equity index, can be calculated in the following way

$$C_t(x,\alpha) = e^{-(r(T)T - r(t)t)} \sum_{(y,\beta)\in\Omega\times V} p((x,\alpha), t; (y,\beta), T) h(e^{-(r(T)-d(T))T} F(y)). \tag{17}$$

The point x from Ω is chosen in such a way so that the equation $F(x) = e^{(r(t)-d(t))t} S_t$ holds, where S_t is the index level at the current time t, and α in V corresponds to the volatility regime we are in at time t.

Our next task is to find the hedge-ratios for the derivative C_0 within our model. It emerges that this is a very simple matter which is not at all computationally demanding because all the hard numerical work has already been done by the pricing algorithm. The delta and the gamma of C_0 are defined using symmetric differences in the following way

$$\begin{aligned} \Delta(x,\alpha) &:= \frac{C_0(x+1,\alpha) - C_0(x-1,\alpha)}{F(x+1) - F(x-1)} \\ \Gamma(x,\alpha) &:= 4\frac{C_0(x+1,\alpha) + C_0(x-1,\alpha) - 2C_0(x,\alpha)}{(F(x+1) - F(x-1))^2}, \end{aligned}$$

where x is the lattice point in Ω that corresponds to the spot level of the index at time 0 (note that at time 0 the spot is the same as the forward). Notice that calculating $C_t(x+1,\alpha)$, or any other value of the option $C_t(y,\alpha)$ with a starting point y different from x, requires no further diagonalization because the pdf we need is given by a different row of the matrix U_s, which has

[18] Notice that the Markov generator $\mathcal{L}$ in this equation acts on the coordinate functions of U_s as a function of the variable (x,α).

already been calculated during the pricing of the original contract $C_t(x,\alpha)$. In fact if one requires the entire delta and gamma profiles of the derivative C_t (like the ones plotted in figures 4 and 5 for call options struck at 100), the most efficient way of doing it is to use a general matrix-vector multiplication routine `gbmv` from LAPACK and then piece together the corresponding Greeks using the above formulae.

A similar procedure can be applied to obtain the vega of the contract C_t. Again we can define it using a symmetric difference (in the stochastic volatility domain V) with a formula

$$\nu(x,\alpha) := \frac{C_0(x,\alpha+1) - C_0(x,\alpha-1)}{\sigma_{\alpha+1} - \sigma_{\alpha-1}},$$

that needs to be suitably amended in case we are in the volatility regime on the boundary of the domain V. The parameters σ_α, for α in the domain V, are the base volatilities in the CEV processes defined in subsection 3.1. Their values can be found in table 1. It is clear that the same computational technique which yielded delta and gamma profiles can be used to find the vega profile of the derivative C_t (see figure 6 for the vega profile of call options with different maturities struck at 100).

Our final task is to find an algorithm for pricing forward starting options. Recall from subsection 2.3 that the payoff of a forward-start is of the form $(S_T - aS_{T'})^+$, where $T' < T$ are a pair of maturities and a is the forward strike. Let us now assume that the current time is 0 and let us reinterpret the formula (17), for $t = T'$, as a forward price of the option contract which starts at the future time T' and expires at time T, conditional on the underlying equity index being at the level $S_{T'} = e^{-(r(T')-d(T'))T'}F(x)$ and the whole system being in the volatility regime α at time T'. The payoff function in (17) is completely general and can therefore depend on the level of the underlying at T'. In order to find the value of the forward-start at time T' we must evaluate a portfolio of $M(N+1)$ call options, one for each element (x,α) of the set $\Omega \times V$, with the corresponding strikes $ae^{-(r(T')-d(T'))T'}F(x)$ using formula (17). The most efficient way of doing this is to collect all the call payoff functions into a matrix and apply a matrix-matrix multiplication routine `gemm` (from LAPACK) on the forward probability kernel (between times T' and T) and the "payoff matrix" we have just created. Notice that both of these matrices are square and have dimension $M(N+1)$, which is hence also true of the product.

Let us now define the function $h : \Omega \times V \to \mathbb{R}$ by requiring that $h(x,\alpha)$ equals the diagonal element of the above product of matrices, which corresponds to the index (x,α). A moment's reflection will show that the function h we have just defined equals the value of the forward-start at time T'. In order to obtain the today's value of the forward starting option, all we need to do is find the current value of the payoff h using expression (17).

We have therefore shown that pricing a forward-start in our framework amounts to one matrix-matrix multiplication and one matrix-vector multiplication. Similarly to the analysis of sensitivities that was carried out for European options earlier in this subsection, we could find delta, gamma, vega and other higher-order Greeks for the forward-starts. However we are not going to pursue this any further; given the basic ideas we have already put forth, these algorithms would be but trivial extensions.

4. Calibration to the vanilla surface

We are now going to calibrate the model for the underlying described in section 3 to the implied volatility surface of the S&P 500 equity index for maturities between 3 months and 10 years (see legend of figure 1 for all market defined maturities used in the calibration) and a broad range of liquid strikes for each maturity. The market data consists of the implied Black-Scholes volatilities for each strike and maturity[19].

[19]Notice that we are using more strikes for longer maturities than for shorter maturities. This is not an inherent requirement but a consequence of the initial structure of the market data we used for calibration of the model.

Our task is to find a set of values for the parameters of the model so that, if we reprice the above options and express their values in terms of the implied Black-Scholes volatilities, we reobtain the market quotes. The main calibration criterion is to minimize the explicit time dependence in order to preserve the correct (i.e. market implied) smile and skew through time. In other words the goal is to have the deterministic time-change function f from subsection 3.4 as close to the identity transformation as possible. The stationarity requirement in the calibration ensures that the forward smile[20] for any maturity still has the desired shape as can be observed in figures 7, 8, 9 and 10.

In order to calibrate our model we select an inhomogeneous grid with $71 = N + 1$ points which is used to span the possible values for the forward rate F_t^{Ω}. We also choose $M = 5$ local volatility regimes in order to capture the correct behaviour of the smile and skew of options on S&P 500.

It is well-known that in general the calibration problem is ill-posed and there is no guarantee that there exists a unique solution. Our calibration strategy was based on an economic interpretation of the market data which gives a specific view of the smile behaviour in the future, conditional upon the underlying index trading inside (or breaking out of) certain "corridors". Market makers often think in those terms when studying the possible scenarios and the effect that they might have on their portfolio. Within our model these levels are set explicitly using the parameter F_α (see table 1). The desired market view (i.e. the shape of the smile conditional on the underlying trading close to that level) can then be expressed by the appropriate choice for the local volatility regime which is favoured by the stochastic vol process as described in subsection 3.3. The stochastic volatility process is given by the Markov generators $\mathcal{G}_\alpha^V$ which are specified below. The short-dated end of the volatility surface (the so-called "gamma regime" of implied volatility) is, in case of the equity index, controlled by down-jump intensity ν_α^-. The long-dated part of the vol surface (i.e. the so-called "vega regime" of implied volatility) is regulated by the CEV parameters σ_α and β_α as well as by the stochastic volatility generators $\mathcal{G}_\alpha^V$.

The set of parameters that was found to work best with the market data for the S&P 500 is reported in table 1. These parameters were discovered by following the guidelines above. No attempts to automate these procedures have been made. It should be noted that, in case of finite-dimensional Markov generators it is easy to compute a clean gradient with respect to the model parameters, which makes algorithms such as Levenberg-Marquardt square minimizing approach easy to apply as long as one has a good starting point.

The S&P 500 index equaled 1195 at the time when the option data was recorded. Throughout the paper a relative value of the index, set at 100, is used for simplicity. The forward price levels $F(x)$, where x is an element of the underlying grid Ω, are also measured on the relative scale. Note also that our starting volatility regime is regime $\alpha = 2$ and we therefore set $\mathbf{F}_2$ to be equal to 100. The values of the parameters β_α are either negative or small and positive in order to keep the skews of the corresponding local volatility regimes decreasing in strike. The choice of the parameters F_α (see table 1) implies that the smaller values of α correspond to the volatility regimes which are more likely if the forward price is trading at a lower level. Therefore it is natural to choose β_α in such a way that it is an increasing function of α in order to account for the recognized leverage effect which stipulates that the implied volatility levels are negatively correlated with the underlying.

5. Path-dependence and the lifting of a Markov generator

We have so far developed a model for the forward rate of the underlying index but have said nothing about its realized variance. In the current section and the two that follow we are going

[20]There is a closed form solution for the value of a forward-start in the Black-Scholes model and the only unknown parameter in that formula is the forward volatility. It is therefore customary to define a forward smile of any model as a function mapping the forward strike to the implied forward volatility which is obtained by inverting the Black-Sholes formula (see subsection 2.3 for the precise definition of a forward strike).

α	σ_α	β_α	$\bar{\sigma}_\alpha$	ν_α^-	ν_α^+	F_α
0	27.0%	-0.50	60%	0.50	0	50.00
1	15.5%	-0.30	60%	0.40	0	90.00
2	12.5%	-0.10	60%	0.35	0	100.00
3	15.5%	0.10	60%	0.25	0	110.00
4	27.0%	0.50	60%	0.04	0	150.00

TABLE 1. Parameters for the local volatility regimes and the jump intensities.

$$\mathcal{G}_0^V = \begin{pmatrix} -4.75 & 4.75 & 0 & 0 & 0 \\ 3.25 & -6.5 & 3.25 & 0 & 0 \\ 0 & 4.75 & -8 & 3.25 & 0 \\ 0 & 0 & 4.75 & -8 & 3.25 \\ 0 & 0 & 0 & 4.75 & -4.75 \end{pmatrix}, \quad \mathcal{G}_4^V = \begin{pmatrix} -4.75 & 4.75 & 0 & 0 & 0 \\ 3.25 & -8 & 4.75 & 0 & 0 \\ 0 & 3.25 & -8 & 4.75 & 0 \\ 0 & 0 & 3.25 & -6.5 & 3.25 \\ 0 & 0 & 0 & 4.75 & -4.75 \end{pmatrix}$$

Markov generators for stochastic volatility ($\alpha = 0, 4$).

$$\mathcal{G}_\alpha^V = \begin{pmatrix} -4.75 & 4.75 & 0 & 0 & 0 \\ 3.25 & -8 & 4.75 & 0 & 0 \\ 0 & 3.25 & -6.5 & 3.25 & 0 \\ 0 & 0 & 4.75 & -8 & 3.25 \\ 0 & 0 & 0 & 4.75 & -4.75 \end{pmatrix}$$

Markov generators for stochastic volatility ($\alpha = 1, 2, 3$).

to build a mechanism that will make it possible to identify the random behaviour of the realized variance of the index. In order to do this we must return to the fundamental theory.

Let $\mathcal{L}$ be a generator for a continuous-time Markov chain F_t defined on a finite state space Ω. In other words the operator $\mathcal{L}$ is given by a square matrix $(\mathcal{L}(x,y))_{x,y\in\Omega}$ which satisfies probability conservation (9) and has non-negative elements off the diagonal. Each element $\mathcal{L}(x,y)$ describes the first order change of the probability that the chain F_t (in this section we drop the superscript Ω to simplify notation) jumps from level $F(x)$ to the level $F(y)$ in the time interval $[t, t+dt]$, where the deterministic function $F : \Omega \to (0,\infty)$ is an injection which defines the image of the process F_t.

Our aim in this section is to extend (i.e. lift) the generator $\mathcal{L}$ to the generator $\widetilde{\mathcal{L}}$, which will describe the dynamics of the lifting (F_t, I_t) of our original process F_t. The component I_t will be a finite-state Markov chain, which we will describe shortly, that contains the relevant path information, up to time t, of the underlying process F_t. The filtration associated with the lifting (F_t, I_t) will NOT contain any information that is not already available in the original filtration of the process F_t. In other words the extension (F_t, I_t) will be adapted to the filtration generated by F_t. On the other hand the Markov generator of the lift (F_t, I_t) will give us the probability kernel for the process I_t, which is what we are ultimately interested in as it contains the probability distribution of the relevant path information.

We will start by describing a continuous process Σ_t which will contain the relevant path-information of the underlying Markov chain and then define a finite-state stochastic process I_t that will be used to model Σ_t. The procedure we are about to describe works for a specific type of path-dependence only. Assume that if at time t the underlying process F_t is at a state $F(x)$ for some $x \in \Omega$, then the change of the value Σ_t, in the infinitesimal time interval dt is of the form

$$d\Sigma_t = Q(x)dt,$$

where the function $Q : \Omega \to \mathbb{R}$, defined in terms of the underlying stochastic process, has two key properties:

- the mapping Q is independent of the path the underlying process follows on the interval $[0, t)$ and
- the value $Q(x)$ depends on the level x at time t and on the distribution of the underlying process in the infinitesimal future time interval dt as given by the Markov generator $\mathcal{L}$.

The first property states that the change in the path-information Σ_t, over a short time period $[t, t + dt]$, does not depend on the path taken by the underlying process up to time t. The second property tells us that the evolution of the path information over the interval $[t, t + dt]$, conditional on the current level of the process F_t, is determined by the future distribution of F_t over the infinitesimal time interval.

It is clear that the realized variance of the process F_t up to time t, can be captured by a random process Σ_t. Indeed, if we define the function Q in the following way

$$Q(x) := \sum_{y \in \Omega} \left(\frac{F(y) - F(x)}{F(x)} \right)^2 \mathcal{L}(x, y) \tag{18}$$

it follows immediately that the above conditions are satisfied. This is because the realized variance of F_t is simply an integral over time of the instantaneous variance of F_t which is given by (18).

The last observation is crucial for all that follows, because it implies that the process Σ_t is uniquely determined by its state-dependent instantaneous drift. In particular we see that the process Σ_t has no volatility and that it has continuous sample paths since it allows a representation as an integral over time.

The fundamental consequence of these facts is the following: a finite-state Markov chain I_t can be used as a model for the process Σ_t if and only if the instantaneous drift of the chain is equal to $Q(x)$, whenever the underlying process F_t is in state $F(x)$, and the instantaneous variance of I_t is equal to zero up to first order. The first requirement clearly follows from the discussion above. The second condition is there to reflect the fact that the process Σ_t is a continuous Itô process which has no volatility term. A non-constant random process on a lattice will always have a non-zero instantaneous variance, but the second condition ensures that this instantaneous variance goes to zero as quickly as the lattice spacing itself. In other words the Markov chain model I_t for Σ_t must exhibit neither diffusion nor jump behaviour.

5.1. **Lifting of the generator for the underlying process.** Recall that the Markov generator $\mathcal{L}(x, y)$ for the underlying process in the forward measure, as described in section 3, is given by (16). Here we have simplified the notation assuming that the variable x (resp. y) now denotes both the lattice value for the spot and the volatility level (the variable x (resp. y) runs over a set with $M(N + 1)$ elements as defined in subsection 3.3).

Let I_t as above denote the Markov chain whose value approximates the realized variance Σ_t of the forward price F_t from time 0 to time t. We shall express I_t as αm_t where m_t is a Poisson process (with non-constant intensity) starting at 0 and gradually jumping up along the grid given by $\{0, \ldots, 2C\}$, where C is an element in $\mathbb{N}$. The constant α controls the spacing of the grid for the realized variance I_t.

We are now going to specify precisely the dynamics of the process m_t which is a fundamental ingredient of our model. As mentioned before, the process m_t will behave as a Poisson process whose intensity is determined by the level of the underlying. In other words the Markov generator, conditional on the underlying process being at the level $F(x)$, is of the form

$$\mathcal{L}^m(x : c, d) := \begin{cases} \frac{1}{\alpha} Q(x) & \text{if } d = (c + 1) \mod (2C + 1); \\ -\frac{1}{\alpha} Q(x) & \text{if } d = c. \end{cases} \tag{19}$$

The variables c and d are elements of the discrete set $\{0, \ldots, 2C\}$ and the function Q is the instantaneous variance of the underlying process as defined in (18). This family of generators

specifies the dynamics of the process I_t with the following instantaneous drift

$$\begin{aligned}\lim_{\Delta t\to 0}\mathbb{E}\left[\frac{I_{t+\Delta t}-I_t}{\Delta t}\Big|F_t=F(x),I_t=I(c)\right] &= \sum_{d=0}^{2C}(I(d)-I(c))\mathcal{L}^m(x:c,d)\\ &= \alpha(c+1-c)\frac{1}{\alpha}Q(x)=Q(x),\end{aligned}\tag{20}$$

for all values $F(x)$ of the underlying process F_t and all integers c which are strictly smaller than $2C$. This implies that the Markov chain I_t has the same instantaneous drift as the actual realized variance. A similar calculation tells us that the instantaneous variance of I_t equals $\alpha Q(x)$. Since α is the spacing of the lattice for the chain I_t, we have just shown that the first two instantaneous moments of I_t match the first two moments of the realized variance Σ_t for all points on the lattice $\{0,\dots,2C\}$ except the last one.

Notice however that the equality (20) breaks down if c equals $2C$, because we have imposed periodic boundary condition for the process I_t. Put differently this means that I_t is in fact a process on a circle rather than on an interval. The latter would be achieved if we had imposed absorbing boundary conditions at $2C$. That would perhaps be a more natural thing to do since the process Σ_t is certainly not periodic. But an absorbing boundary condition would destroy the delicate structure of the spectrum of the lifted generator $\widetilde{\mathcal{L}}$ which is preserved by the periodic boundary condition. It is precisely this structure that makes the periodic nature of I_t a key ingredient of our model, because it allows us to linearize the complexity of the pricing algorithm, as we shall see in section 6. It should be noted that the general philosophy behind either choice of the boundary condition would be the same: the lattice in the calibrated model should be set up in such a way that the process never reaches the boundary value $2C$, because if it does an inevitable loss of information will ensue regardless of the boundary conditions we choose.

We are now finally in a position to define the lifted Markov generator of the process (F_t, I_t) as

$$\widetilde{\mathcal{L}}(x,c;y,d) := \mathcal{L}(x,y)\delta_{c,d}+\mathcal{L}^m(x:c,d)\delta_{x,y}.\tag{21}$$

The structure of the spectrum of the operator $\widetilde{\mathcal{L}}$ will be exploited in section 7 to obtain a pricing algorithm for payoffs which are general functions of the realized variance. The reason for specifying the generator $\mathcal{L}^m(x:c,d)$ by (19) (i.e. insisting on the periodic boundary condition for the process I_t) will become clear in section 6 when we explore the spectral properties of the lifted generator $\widetilde{\mathcal{L}}$.

6. Diagonalization algorithm for partial-circulant matrices

In this section our aim is to generalize a known diagonalization method from linear algebra which will yield a numerically efficient algorithm for obtaining the joint probability distribution of the spot and realized variance at any maturity. Let us start with some well-known concepts.

A matrix $C\in\mathbb{R}^{n\times n}$ is *circulant* if it is of the form

$$C=\begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1}\\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2}\\ c_{n-2} & c_{n-1} & c_0 & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & c_1\\ c_1 & c_2 & \cdots & c_{n-1} & c_0\end{pmatrix},$$

where each row is a cyclic permutation of the row above. The structure of matrix C can also be expressed as

$$C_{ij}=c_{(i-j)\bmod n},$$

where C_{ij} is the entry in the i-th row and j-th column of matrix C. It is clear that any circulant matrix is a Toeplitz[21] operator and in fact circulant matrices are used to approximate general Toeplitz matrices and explain the asymptotic behaviour of their spectra. We will not investigate this idea any further (for more information on the topic see (Bottcher & Silbermann 2006)) since our main interest lies in a different generalization of circulant matrices, namely that of partial-circulant matrices, which will be defined in subsection 6.2. Before doing that we are going to recall some of the known properties of circulant matrices.

6.1. **Eigenvalues and eigenvectors of circulant matrices.** Let C denote a circulant matrix of dimension n as defined above. The eigenvalue λ and the eigenvector $y \in \mathbb{R}^n$ are the solutions of the equation $Cy = \lambda y$, which is equivalent to the system of n linear difference equations with constant coefficients:

$$\begin{aligned} \sum_{k=o}^{n-1} c_k y_k &= \lambda y_0 \quad \text{and} \\ \sum_{k=0}^{j-1} c_{n-j+k} y_k + \sum_{k=j}^{n-1} c_{k-j} y_k &= \lambda y_j, \quad \text{for} \quad j \in \{1, \ldots, n-1\}. \end{aligned}$$

The variables y_k, for $k \in \{0, \ldots, n-1\}$, in these equations are simply the coordinates of the eigenvector y. Such systems are routinely solved by guessing the solution and proving that it is correct (see appendix 1 in (Grimmett & Stirzaker 2001)). The solution in this case is of the form

$$\lambda = \sum_{k=0}^{n-1} c_k z^k \quad \text{and} \quad y_j = \frac{z^j}{\sqrt{n}} \quad \text{for} \quad j \in \{0, \ldots, n-1\},$$

where z is a complex number which satisfies $z^n = 1$. This implies that the eigenvalue-eigenvector pairs of matrix C are parameterized by the n-th roots of unity which are of the form $z_r = \exp(-2\pi i r/n)$, where the index r lies in $\{0, \ldots, n-1\}$ and i is the imaginary unit. Therefore the j-th coordinate of the r-th eigenvector, together with the corresponding eigenvalue, can be expressed as

$$y_j^{(r)} = \frac{1}{\sqrt{n}} e^{-i\frac{2\pi}{n} rj} \quad \text{and} \tag{22}$$

$$\lambda_r = \sum_{k=0}^{n-1} c_k e^{-i\frac{2\pi}{n} rk} \quad \text{for} \quad r, j \in \{0, \ldots, n-1\}. \tag{23}$$

This representation is extremely useful because it allows us to deduce a number of fundamental facts about circulant matrices. Let us start with the eigenvectors. It is obvious that if we put all n vectors $y^{(r)}$ side by side into a matrix, the determinant of the linear operator obtained is the Vandermonde determinant, which is non-zero since its parameters are the n distinct solutions of the equation $z^n = 1$. This implies that matrix C can be diagonalized and that all its eigenvectors are of the form (22).

Another key property of circulant matrices is that they can all be diagonalized using the same set of eigenvectors. This follows directly from (22) since the expression for the vectors $y^{(r)}$ are clearly independent of matrix C.

Expression (23) tells us that the r-th eigenvalue of C equals the value (at the point r) of the discrete Fourier transform (DFT) of the sequence $(c_j)_{j=0,\ldots,n-1}$. We can therefore recover the sequence (c_j) from the spectrum $(\lambda_r)_{r=0,\ldots,n-1}$ of C using the inverse discrete Fourier transform. Even though this is a very well-known and celebrated fact, we will now present a short proof for it, because the argument itself sheds light on the behaviour of circulant matrices.

[21]For definition see for example (Bottcher & Silbermann 2006). Toeplitz operators arise in many contexts in theory and practice and therefore constitute one of the most important classes of non-self-adjoint operators. They provide a setting for a fruitful interplay between operator theory, complex analysis and Banach algebras.

Note that for any index $k \in \mathbb{Z}$, such that $(k \bmod n)$ is different from zero, we obtain

$$\sum_{r=0}^{n-1} e^{i\frac{2\pi}{n}rk} = \frac{1 - e^{i\frac{2\pi}{n}kn}}{1 - e^{i\frac{2\pi}{n}k}} = 0, \tag{24}$$

by summing a finite geometric series. In particular this implies that the above sum for any $k \in \mathbb{Z}$ equals $n\delta_{1,k \bmod n}$, where δ is the Kronecker delta which takes value 1 at zero and value 0 everywhere else. The inversion formula for the DFT is now an easy consequence

$$\begin{aligned} \frac{1}{n}\sum_{r=0}^{n-1} \lambda_r e^{i\frac{2\pi}{n}rl} &= \frac{1}{n}\sum_{r=0}^{n-1}\sum_{k=0}^{n-1}\left(c_k e^{-i\frac{2\pi}{n}rk}\right) e^{i\frac{2\pi}{n}rl} \\ &= \sum_{k=0}^{n-1} c_k \frac{1}{n}\sum_{r=0}^{n-1} e^{i\frac{2\pi}{n}r(l-k)} = c_l, \end{aligned}$$

for any l in $\{0, \ldots, n-1\}$. Before proceeding we should note that the argument we have just outlined implies that a circulant matrix is uniquely[22] determined by its spectrum.

Another consequence of the extraordinary identity (24) is that for any pair of distinct indices k and r in $\{0, \ldots, n-1\}$, the corresponding eigenvectors $y^{(k)}$ and $y^{(r)}$ are perpendicular to each other. Since we have chosen the vectors $y^{(r)}$ in (22) so that their norm is one, the set of all eigenvectors of a circulant matrix is an orthonormal basis of the vector space $\mathbb{C}^n$.

Let A be another circulant matrix given by the sequence $(a_k)_{k=0,\ldots,n-1}$ with the spectrum $(\alpha_r)_{r=0,\ldots,n-1}$. Since A and C can be diagonalized simultaneously using the basis $\{y^{(r)};\ r = 0, \ldots, n-1\}$, it follows that the product AC is also diagonal in this basis and that its eigenvalues are of the form $\alpha_r\lambda_r$. Therefore AC is a circulant matrix whose first row equals the convolution[23] of the sequences (a_k) and (c_k). The diagonal representation also implies that the matrices A and C commute. Finally note that the sum $A + C$ is also a circulant matrix.

6.2. **Partial-circulant matrices.** We are now going to define a class of matrices, that will include the Markov generator given by (21), which can be diagonalized by the semi-analytic algorithm from subsection 6.4.

Let A be a linear operator represented by a matrix in $\mathbb{R}^{m\times m}$ and let $B^{(k)}$, for $k = 0, \ldots, m-1$, be a family of n-dimensional matrices with the following property: there exists an invertible matrix $U \in \mathbb{C}^{n\times n}$ such that

$$U^{-1}B^{(k)}U = \Lambda^{(k)}, \quad \text{for all} \quad k \in \{0, \ldots, m-1\},$$

where $\Lambda^{(k)}$ is a diagonal matrix in $\mathbb{C}^{n\times n}$. In other words this condition stipulates that the family of matrices $B^{(k)}$ can be simultaneously diagonalized by the transformation U. Therefore the columns of matrix U are eigenvectors of $B^{(k)}$ for all k between 0 and $m-1$.

Let us now define a large linear operator $\widetilde{A}$, acting on a vector space of dimension mn, in the following way. Clearly matrix $\widetilde{A}$ can be decomposed naturally into m^2 blocks of size $n \times n$. Let $\widetilde{A}_{i,j}$ denote an $n \times n$ matrix which represents the block in the i-th row and j-th column of this decomposition. We now define the operator $\widetilde{A}$ as

$$\widetilde{A}_{ii} := B^{(i)} + A_{ii}\mathbb{I}_{\mathbb{R}^n} \quad \text{and} \tag{25}$$

$$\widetilde{A}_{ij} := A_{ij}\mathbb{I}_{\mathbb{R}^n}, \quad \text{for all} \quad i, j \in \{1, \ldots, m\} \quad \text{such that} \quad i \neq j. \tag{26}$$

The real numbers A_{ij} are the entries of matrix A and $\mathbb{I}_{\mathbb{R}^n}$ is the identity operator on $\mathbb{R}^n$. We may now state our main definition.

Definition. A matrix is termed *partial-circulant* if it admits a structural decomposition as in (25) and (26) for any matrix $A \in \mathbb{R}^{m\times m}$ and a family of n-dimensional circulant matrices $B^{(k)}$, for $k = 0, \ldots, m-1$.

[22] Such a statement is untrue even for self-adjoint and unitary operators.

[23] Recall that the DFT of the convolution of two sequences equals the product of the DFTs of each of the sequences.

The concept of a partial-circulant matrix is well-defined because, as we have seen in subsection 6.1, any family of circulant matrices can be diagonalized by a unitary transformation whose columns consist of vectors $y^{(r)}$, for r between 0 and $n-1$ (see equation (22)).

The operator $\widetilde{A}$ is very big indeed. The typical values that are of interest to us for the dimensions m and n are 70 and 600 respectively. This implies that matrix $\widetilde{A}$ contains $(70\cdot 600)^2$, i.e. more than one billion, entries. This means that even storing $\widetilde{A}$ on a computer requires about 10 Gb of memory.

Our task is to find the spectrum of the operator $\widetilde{A}$. Given its size and the fact that it is not a sparse matrix, this problem at first sight appears not to be tractable. But the structure of matrix $\widetilde{A}$, combined with the ubiquitous idea of invariant subspaces of linear operators, will yield the solution. We will describe the diagonalization algorithm for partial-circulant matrices in subsection 6.4. Before we do this we need to recall the basic properties of invariant subspaces.

6.3. **Invariant subspaces of linear operators.** Let $A : V \to V$ be a linear operator on a finite-dimensional vector space V. By definition a subspace X of V is an *invariant subspace* of the operator A if and only if $AX \subseteq X$. Note that the set AX is a subspace of V. It is clear from the definition that vector spaces $\{0\}$, V, AV and $\ker(A)$ are all invariant subspaces of the operator A. Another trivial example is the space of all eigenvectors of A that belong to an eigenvalue λ.

It is the non-trivial examples however that make this concept so powerful. If we can find two invariant subspaces X_1 and X_2 of V for the operator A, such that $X_1 \cap X_2 = \{0\}$ and $\dim X_1 + \dim X_2 = \dim V$ (i.e. $V = X_1 \oplus X_2$), then in the appropriate basis the matrix representing the operator A takes the form

$$D = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

where A_1 (resp. A_2) is the matrix acting on the subspace X_1 (resp. X_2). The zeros in the above expression represent trivial linear operators that map the subspace X_1 into the origin of the subspace X_2 and vice versa.

The advantage of this structural decomposition of the original operator A is clear because it reduces the dimensionality of the problem. The spectral decomposition (i.e. the eigenvalues and eigenvectors) of A can now be obtained from the spectral decomposition of two smaller operators A_1 and A_2. *Block-diagonalization* consists of finding the transition matrix F (i.e. the appropriate coordinate change) that will transform the original matrix A into block-diagonal form given by matrix D above:

$$F^{-1}AF = D.$$

6.4. **Algorithm for block-diagonalization.** Let $\widetilde{A}$ be the linear operator defined in (25) and (26) which acts on the vector space $\mathbb{C}^{mn}$. We are now going to describe the block-diagonalization algorithm for $\widetilde{A}$. In other words we are going to find invariant subspaces V_j of the operator $\widetilde{A}$ (where j ranges between 1 and n), such that $\mathbb{C}^{mn} = V_1 \oplus \cdots \oplus V_n$, and a transition matrix $F \in \mathbb{C}^{mn\times mn}$, such that the only non-zero $n \times n$ blocks of matrix $F^{-1}\widetilde{A}F$ are the diagonal ones.

Recall that, by definition of $\widetilde{A}$, there exists a matrix $U \in \mathbb{C}^{n\times n}$ consisting of eigenvectors for the matrices $B^{(k)}$. Put differently the columns $u_j \in \mathbb{C}^n$, for $j = 1, \ldots, n$, of U satisfy the identity

$$B^{(k)}u_j = \lambda_j^{(k)}u_j \quad \text{for} \quad k \in \{0, \ldots, m-1\}.$$

Now fix any index j between 1 and n and define vectors $v_i^{(j)} \in \mathbb{C}^{mn}$, where $i = 1, \ldots, m$, as follows:

$$v_i^{(j)} := (\underbrace{0, \ldots, 0}_{(i-1)n}, u_j', \underbrace{0, \ldots, 0}_{(m-i)n})', \tag{27}$$

where u'_j is a row of n complex numbers obtained by transposing and conjugating the vector u_j. We can now define the subspace V_j of $\mathbb{C}^{mn}$ as the linear span of vectors $v_i^{(j)}$. It is clear that the intersection of subspaces V_j and V_k is trivial for any two distinct indices $j, k \in \{1, \dots, n\}$. This is because the eigenvectors u_j and u_k are linearly independent in $\mathbb{C}^n$ since, by assumption, matrix U is invertible. It follows directly from the definition that the dimension of V_j is m. Since there are exactly n subspaces V_j, we obtain the decomposition $\mathbb{C}^{mn} = V_1 \oplus \cdots \oplus V_n$.

If we manage to show that each space V_j is an invariant subspace for the operator $\widetilde{A}$, we will be able to conclude that $\widetilde{A}$ can be expressed in the block-diagonal form as described in subsection 6.3. Since the subspace V_j is defined as a linear span of a set of vectors $\{v_i^{(j)};\ i = 1, \dots, m\}$, the invariance property $\widetilde{A}V_j \subseteq V_j$ will follow if we demonstrate that the vector $\widetilde{A}v_i^{(j)}$ is in V_j for all $i = 1, \dots, m$. By definition of $\widetilde{A}$ ((25) and (26)) it immediately follows that

$$\widetilde{A}v_i^{(j)} = \sum_{k=1}^{m} A_{ik} v_k^{(j)} + \lambda_j^{(i-1)} v_i^{(j)} \quad \text{for} \quad i \in \{1, \dots, m\}, \tag{28}$$

where $\lambda_j^{(i-1)}$ is the eigenvalue of matrix $B^{(i-1)}$ that corresponds to the eigenvector u_j and the real numbers A_{ki} are the entries of matrix $A \in \mathbb{R}^{m\times m}$. Identity (28) implies that each subspace V_j is an invariant subspace for $\widetilde{A}$. Furthermore, if we define a matrix $F \in \mathbb{C}^{mn\times mn}$ in the following way

$$F := \left(v_1^{(1)}, \dots, v_m^{(1)}, v_1^{(2)}, \dots, v_m^{(2)}, \dots, v_1^{(n)}, \dots, v_m^{(n)}\right), \tag{29}$$

then matrix $D = F^{-1}\widetilde{A}F$ is block-diagonal. In other words if we decompose D into n^2 matrices D_{ij} of size $m \times m$, then the following formula holds

$$D_{ij} = \delta_{ij}(A + \Theta^{(j)}) \quad \text{for} \quad i, j \in \{1, \dots, n\}, \tag{30}$$

where $\Theta^{(j)}$ is a diagonal matrix in $\mathbb{C}^{m\times m}$ with its i-th diagonal element equal to $\lambda_j^{(i-1)}$. As usual the symbol δ_{ij} denotes the Kronecker delta function.

Expression (30) gives us the block-diagonal representation of the operator $\widetilde{A}$. Notice that the diagonal elements of matrix $\Theta^{(j)}$ are precisely the eigenvalues of matrices $B^{(i)}$, for $i = 0, \dots, m-1$, that correspond to the eigenvector u_j.

The algorithm to block-diagonalize the operator $\widetilde{A}$, defined by matrices $A \in \mathbb{R}^{m\times m}$ and $B^{(k)} \in \mathbb{R}^{n\times n}$ (see (25) and (26)), can now be described as follows:

(I) Find matrix $U \in \mathbb{C}^{n\times n}$ whose columns are the common eigenvectors u_j, for $j \in \{1, \dots, n\}$, of the family $B^{(k)}$.
(II) Construct the transition matrix F using the columns of matrix U as described in (27) and (29).
(III) Find the eigenvalues $\lambda_j^{(k)}$ which satisfy $B^{(k)}u_j = \lambda_j^{(k)}u_j$ for $k \in \{0, \dots, m-1\}$ and $j \in \{1, \dots, n\}$.
(IV) Construct diagonal matrices $\Theta^{(j)} \in \mathbb{C}^{m\times m}$, for all $j \in \{1, \dots, n\}$, given by $\Theta_{ik}^{(j)} = \delta_{ik}\lambda_j^{(i-1)}$, where the indices i, k run over the set $\{1, \dots, m\}$.
(V) Construct the block-diagonal representative D for the operator $\widetilde{A}$ as described in (30).

Our main task is to find the spectrum of matrix $\widetilde{A}$. Notice that, since the spectrum of $\widetilde{A}$ is a union of the spectra of $A + \Theta^{(j)}$, this algorithm has reduced the problem of diagonalizing an $nm \times nm$ matrix to finding the spectra of n matrices of size $m \times m$. The algorithm provides a key step for our pricing method because it enables us to model the behaviour of the realized variance by increasing the numerical complexity only linearly .

We should also note that in the case of the lifted Markov generator $\widetilde{\mathcal{L}}$ in (21), matrix A is the generator $\mathcal{L}$ of the underlying process while the family $B^{(k)}$ consists of circulant matrices.

In other words the operator $\widetilde{\mathcal{L}}$ is given by a partial-circulant matrix. It therefore follows from the discussion in subsection 6.1 that the columns of the corresponding transition matrix F are pairwise orthogonal and that the entries of matrix $\widetilde{\mathcal{L}}F$ are the values of a partial[24] discrete Fourier transform of the rows of $\widetilde{\mathcal{L}}$. This simple observation will be useful when calculating the probability kernel of the lifted generator in the next section.

7. Pricing of derivatives on realized variance

Let Σ_T be the realized variance over the time interval $[0, T]$, expressed in annual terms, as defined in (1). In this section we are going to find pricing formulae, analogous to those given in subsection 3.5, for general payoffs that depend on the annualized realized variance Σ_T.

Our first task is to find the probability kernel of the lifted process (F_t, I_t) which was defined in section 5. Recall that the Markov chain $I_t = \alpha m_t$, which is used to model the realized variance, is specified in terms of a translation invariant process m_t, whose domain is $\Psi = \{0, \dots, 2C\}$, and some positive constant α which determines the lattice spacing for the domain of I_t. The value $C \in \mathbb{N}$ specifies the size of the lattice for the realized variance and has to be large enough so that the process I_t, starting from zero, does not transverse the entire lattice. This is a very important technical requirement as it ensures that there is no probability leakage in the model (which is theoretically possible since we are using periodic boundary conditions for the process I_t). The dynamics of the chain m_t are given by the Markov generator in (19).

Recall also that the process I_t records the total realized variance up to time t, which implies that the annualized realized variance Σ_T that interests us, will be described by $\frac{1}{T} I_T$ where the time horizon T is expressed in years. The key ingredient in the calculation of the probability distribution function of the process (F_t, I_t) is the block-diagonalization algorithm from subsection 6.4. We are now going to apply it to the generator (21) in order to find the joint pdf of the lifted process.

7.1. Probability kernel of the lift $(\mathbf{F_t}, \mathbf{I_t})$.

We saw at the end of section 6 that the generator $\widetilde{\mathcal{L}}(x, \beta, c; y, \gamma, d)$ of the process (F_t, I_t), given by (21), is a square partial-circulant matrix acting on a vector space of dimension $M(N+1)(2C+1)$. The coordinates $(x, \beta, c), (y, \gamma, d)$ of the matrix $\widetilde{\mathcal{L}}$ (i.e. the lattice points of the process) lie in the set $\Omega \times V \times \Psi$, where Ω is the grid for the underlying forward rate and the set V contains all volatility regimes of the model (see section 3 for a precise definition). Notice that the circulant matrices $\mathcal{L}^m$ from (19), used in the definition of $\widetilde{\mathcal{L}}$, are very simple because the only non-zero elements are on the diagonal, just above the diagonal and the element in the bottom left corner of the matrix. In other words if we interpret the matrix $\mathcal{L}^m$, associated with the lattice point (x, β), in terms of the definition of a circulant matrix given at the beginning of section 6, we see that

$$c_1 = -c_0 = \frac{Q(x, \beta)}{\alpha},$$

where the function $Q(x, \beta)$ is the instantaneous variance as defined by (18) and the constant α is the lattice spacing of the domain of I_t. All other elements c_j are equal to zero.

It is therefore clear that, using the expression (23) for the eigenvalues of circulant matrices, equation (30) can be reinterpreted as

$$\mathcal{L}_k(x, \beta; y, \gamma) := \mathcal{L}(x, \beta; y, \gamma) + \delta_{(x,\beta),(y,\gamma)} \left(e^{-ip_k} - 1\right) \frac{Q(x, \beta)}{\alpha},$$

where $\mathcal{L}_k$ is the k-th block in the block-diagonal decomposition of $\widetilde{\mathcal{L}}$ and the value of p_k is given by the expression

$$p_k := \frac{2\pi}{2C+1} k. \tag{31}$$

[24] Since each row of matrix $\widetilde{\mathcal{L}}$ is naturally described by two variables, namely the value of the underlying and the value of the realized variance, partial DFT is by definition a DFT acting on the second variable.

The index k in these expressions runs from 0 to $2C$. Notice that the matrices $\mathcal{L}_k$ differ from the Markov generator $\mathcal{L}$ only along the diagonal. We are now in a position to state the key theorem that will allow us to find the probability kernel of the lifted process.

Theorem 7.1. *Let $\widetilde{\mathcal{L}}$ be the Markov generator of the stochastic process (F_t, I_t) as described in section 5 and let $\phi : \mathbb{C} \to \mathbb{C}$ be a holomorphic function. Then the following equality holds*

$$\phi(\widetilde{\mathcal{L}})(x,\beta,c;y,\gamma,d) = \frac{1}{2C+1}\sum_{k=0}^{2C} e^{-ip_k(c-d)}\phi(\mathcal{L}_k)(x,\beta;y,\gamma),$$

where $\mathcal{L}_k$ is the operator defined above, p_k is given by (31) and $(x,\beta,c),(y,\gamma,d)$ are elements of $\Omega \times V \times \Psi$.

Before embarking on the proof of this theorem we may summarize as follows: if a linear operator A can be block-diagonalized by a discrete Fourier transform (cf. last paragraph of subsection 6.4), then so can any operator $\phi(A)$ where ϕ is a holomorphic function defined on the entire complex plane. Note that the assumptions on function ϕ and the linear operator $\widetilde{\mathcal{L}}$ in theorem 7.1 are too stringent and in fact the theorem holds in much greater generality. For our purposes however the setting described in the theorem is sufficient as it applies directly to the model. We will therefore only give a proof of the restricted case stated above.

Proof. Let us start by recalling that any holomorphic function defined on $\mathbb{C}$ has a Taylor expansion around zero that converges everywhere. We can therefore define $\phi(A)$, via expression (13), for any linear operator A on a finite-dimensional vector space. It also follows from the fact that ϕ has a Taylor expansion that any invariant subspace (see subsection 6.3 for definition) of A is also an invariant subspace of $\phi(A)$. In particular if A has a block-diagonal decomposition in the sense of subsection 6.4, then the matrix $\phi(A)$ also has one. Moreover if B is a block in A then $\phi(B)$ must be a block in $\phi(A)$.

We know that the Markov generator $\widetilde{\mathcal{L}}$ can be expressed as $\widetilde{\mathcal{L}} = FDF^{-1}$, where D is a block-diagonal matrix of the form

$$D = \begin{pmatrix} \mathcal{L}_0 & 0 & \cdots & 0 \\ 0 & \mathcal{L}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{L}_{2C} \end{pmatrix},$$

and the transition matrix F is given by (29). We have just seen that $\phi(D)$ must therefore also be in block-diagonal form:

$$\phi(D) = \begin{pmatrix} \phi(\mathcal{L}_0) & 0 & \cdots & 0 \\ 0 & \phi(\mathcal{L}_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi(\mathcal{L}_{2C}) \end{pmatrix}.$$

It is clear from definition (13) and the power series expansion of ϕ that $\phi(\widetilde{\mathcal{L}}) = F\phi(D)F^{-1}$. Since matrix F is defined using the eigenvectors of circulant matrices, it follows immediately that the inverse F^{-1} can be obtained by transposing F and conjugating each of its elements. Note that the dimension of our circulant matrices is $2C+1$ and express $\phi(\widetilde{\mathcal{L}})(x,\beta,c;y,\gamma,d) = \langle u, \phi(D)v\rangle$ as a real inner product of two vectors u and $\phi(D)v$, where u equals the (x,β,c)-row of the matrix F and v is the (y,γ,d)-column of F^{-1} (i.e. the conjugated (y,γ,d)-row of F).

It follows from the definition of F and the above expression for $\phi(D)$ that the non-zero coordinates of the vector u are of the form

$$\frac{1}{\sqrt{2C+1}} e^{-ip_k c}$$

for all $k \in \Psi$ and that the corresponding coordinates of $\phi(D)v$ are

$$\frac{1}{\sqrt{2C+1}} e^{ip_k d} \phi(\mathcal{L}_k)(x, \beta; y, \gamma).$$

The equality in the theorem now follows directly from the expressions for the coordinates of the vectors u and $\phi(D)v$ and the definition of the real inner product. This concludes the proof of the theorem. □

We can now find the full probability kernel of the process (F_t, I_t) by applying theorem 7.1 and then using the same procedure as in subsection 3.5. More explicitly, for any pair of calendar times t and T, such that $t < T$, we have

$$\begin{aligned} p((x,\beta,c),t;(y,\gamma,d),T) &= e^{(f(T)-f(t))\tilde{\mathcal{L}}}(x,\beta,c;y,\gamma,d) \\ &= \frac{1}{2C+1}\sum_{k=0}^{2C} e^{-ip_k(c-d)} e^{(f(T)-f(t))\mathcal{L}_k}(x,\beta;y,\gamma) \\ &= \frac{1}{2C+1}\sum_{k=0}^{2C}\sum_{n=1}^{M(N+1)} e^{\lambda_n^k (f(T)-f(t))} e^{-ip_k(c-d)} u_n^k(x,\beta) v_n^k(y,\gamma), \end{aligned} \tag{32}$$

where λ_n^k are the eigenvalues and u_n^k are the eigenvectors of $\mathcal{L}_k$. As usual we denote by v_n^k the columns of the matrix U_k^{-1}, where U_k consists of all the eigenvectors of $\mathcal{L}_k$. Function f in the above formula is the deterministic time-change for the underlying model which was introduced in subsection 3.4.

Formula (32) is our key result because it allows us to price any derivative which depends jointly on the realized variance of the underlying index and the index itself at any time horizon T. In subsection 7.2 we find explicit formulae for the values of such derivatives using the joint probability distribution function (32).

Figures 13, 14, 15, 16, 17 and 18 contain the graphs of joint distribution functions of the spot level and the annualized realized variance for maturities between 6 months and 5 years. The marginal distributions of the annualized realized variance, obtained from the above joint distributions by integration in the dimension of the spot value of the index, are shown in figure 11.

The parameters in the model that influence the dynamics of the realized variance I_t are the number of lattice points C and the lattice spacing α. Table 2 contains their values for different maturities. It should be noted that the numerical complexity of the algorithm used to obtain

	6m	1y	2y	3y	4y	5y
C	80	80	100	90	100	110
α	0.0002	0.0004	0.0009	0.0017	0.0021	0.025

TABLE 2. Parameters specifying the geometry of the lattice for the realized variance process I_t.

the pdf of the joint process (F_t, I_t) grows linearly with C. As mentioned earlier, it is crucial that the value of C be chosen large enough with respect to the spacing α so that the process cannot get to the other side of the lattice with positive probability. Having arrived at the pdf in (32), our next task is to find the value of any derivative whose payoff depends on the realized variance.

7.2. **Pricing derivatives on the realized variance.** Let us assume that we are given a general payoff $h(\Sigma_{T-t})$ that depends on the annualized realized variance Σ_{T-t} between the current time t and some future expiry date T. The current price of a derivative with this payoff within our model can be computed directly using the joint probability distribution function (32) in the

following way

$$C_t(x,\beta) = e^{-(r(T)T - r(t)t)} \sum_{d\in\Psi} \left(\sum_{(y,\gamma)\in\Omega\times V} p((x,\beta,0),t;(y,\gamma,d),T) \right) h\left(\frac{\alpha d}{T-t}\right).$$

The sum in the brackets is the marginal distribution of the process I_t at time $T-t$. Notice that the realized variance process must always start from 0 at the inception of the contract. The factor $\frac{1}{T-t}$ normalizes the value of I_{T-t} so that it is expressed in annual terms. As before, the constant α is the lattice spacing for the realized variance.

The point x from Ω is chosen in such a way that the equation $F(x) = e^{(r(t)-d(t))t}S_t$ holds, where S_t is the index level at the current time t, and β in V corresponds to the volatility regime at time t.

Notice that because in the valuation formula above there is no restriction on the payoff function h, the contract C_t can be anything from a variance swaption to a volatility swap (or swaption) and can be priced almost instantaneously using the calibrated model.

8. Some numerical results

Let us now use the calibrated model together with the pricing and hedging algorithms described in subsection 3.5 and sections 6 and 7 to perform some numerical experiments and consistency checks for vanilla options, forward-starts and variance swaps.

8.1. **Profiles of the Greeks.** In order to test the pricing methodology of the model for the underlying forward rate we pick a strip of call options with the same notional but with varying maturities, all struck at the current spot level. Since the entire framework for the underlying is expressed in relative terms with respect to the current value of the index, the strike used for this strip of options is 100.

Because we are interested in the behaviour of our pricing algorithm in changing market conditions, we are going to study the properties of the first-order sensitivities to the market parameters of our options with respect to all possible spot levels. In other words we are going to calculate delta, gamma and vega, using the definitions given in subsection 3.5, as functions of the current spot level.

This task does not pose any additional numerical difficulties because all it requires is the knowledge of the probability distribution functions of the underlying at the relevant maturities conditional upon the starting level, which can be any of the points in the grid. But if we have already priced a single option, then these pdfs are available to us without any further numerical efforts. This is because, when pricing an option, the algorithm described in subsection 3.5 calculates the entire probability kernel for each starting point in the grid, even though the option pricing formula only requires one row of the final result. In this situation we require all the rows of the probability kernel so as to obtain the prices of our option, conditional upon different spot levels, by applying the matrix of the kernel to the vector whose coordinates are values of the payoff calculated at all lattice points. As defined in subsection 3.5, the Greeks are linear combinations of the coordinates of the final result of the calculation.

Figure 4 contains the delta profile of the call options for different maturities. Figures 5 and 6 give the gamma and the vega profiles of the call options in the strip. As expected, an owner of a vanilla option is both long gamma and long vega. The shapes of the graphs in figures 5 and 6 also confirm that, according to our model, the at-the-money options have the largest possible vega and gamma for any given maturity. A cursory inspection of the scales of gamma and vega for options with the same notional indicates that in our model some calender spreads[25] can be simultaneously long vega and short gamma or vice versa.

[25]A calendar spread is a structure defined by going long one call option and going short another call option of the same strike but different maturity.

8.2. **Forward smile.** In subsection 2.3 we defined the concept of forward volatility using the Black-Scholes formula for the forward starting options. One of the parameters in this formula is a forward strike. Any model for the underlying process defines a functional relationship between forward strikes and implied forward volatilities using the Black-Scholes pricing formula for forward-starts, in much the same way as it defines the implied volatility as a function of strike. This functional relationship is know as the *forward smile*.

The reason why the forward smile is so important lies in the fact that it determines the conditional behaviour of the process. It is well-known that knowing all the vanilla prices (i.e. the entire implied volatility surface) is not enough to price path-dependent exotic options. In terms of stochastic processes this statement can be expressed by saying that knowing the probability distributions of the underlying for all maturities does not determine the process uniquely, not even up to sets of probability zero. The forward smile contains the information about the two-dimensional distributions of the underlying process. In other words one can have two different models that are perfectly calibrated to the implied volatility surface but which assign completely different values to the forward starting options.

Market participants can express their views on the two-dimensional distributions of the underlying process by setting the prices of the forward starting options accordingly. It is therefore of utmost importance for any model used for pricing path-dependent derivatives to have the implied forward smiles close to the ones implied by the market. Unfortunately we did not have the market implied forward volatilities that were prevailing at the time when the S&P 500 vanilla options data was recorded. Figures 7, 8, 9 and 10 contain the forward smiles implied by our model for maturities between 3 months and 2 years. Since we have no market data to compare them with, we can only say that the qualitative nature of the implied forward smiles is as expected in the following sense: for a fixed time T' the forward smiles are flattening with increasing T (for definition of times T' and T see subsection 2.3) and for a fixed difference $T-T'$ the shapes of the forward smiles look similar when compared across maturities T'. It should be noted that the latter point exemplifies the stationary nature of the underlying model and shows that we did not have to use extreme values of the model parameters to calibrate it to the entire implied volatility surface, because the two-dimensional distributions of the underlying process have some of the necessary features which are expected by the market participants.

We should also note that a statistical comparison of the forward volatility smiles implied by the model and by the market can be carried out easily because, as was mentioned at the end of subsection 3.5, pricing a forward starting option consists of two consecutive linear algebra operations which require little computing time.

8.3. **Probability distribution for the implied volatility index.** One of our goals in this paper has been to describe the random evolution of VIX through time. In figure 20 we plot the probability distribution functions of the volatility index, as defined at the end of subsection 2.3. It is clear from these graphs that the future behaviour of the portfolio of options defining VIX is model-dependent. It is not hard to see that the peaks of the distributions in figure 20 are close to the values of the model parameter σ_α (see section 4 for definition) which is used to define different local volatility regimes. This can be explained by recalling that the stochasticity of volatility in our model does not (on average) influence the behaviour of the underlying process in time horizons as short as one month. Therefore the process is very likely to stay within the same local volatility regime for a month after it hits any of the maturities in figure 20. This claim is substantiated by the graphs in figure 21 which exhibit much more regular behaviour (i.e. there are no local maxima) if we monitor the same portfolio as the one in the definition of the volatility index but over a longer time period.

Since we can calculate explicitly distributions of VIX for any maturity (like those in figure 20), pricing a European option on the VIX in our framework amounts to summing the values of the payoff against the pdf in the same way as was done for European payoffs on the underlying index in (17) of subsection 3.5.

8.4. **Distribution of the realized variance.** Most of the modelling exercise presented so far has been directed towards finding the probability distribution function for the annualized realized variance of the underlying process. Figure 11 contains the pdfs for the realized variance for maturities between 6 months and 5 years as implied by the model that was calibrated to the market data for the S&P 500. The corresponding term structure of the fair values of variance swaps for these maturities is given in figure 19. This term structure is obtained by calculating the expectation (in the risk-neutral measure) of the probability distribution functions of figure 11.

Figure 19 also contains the current value of the logarithmic payoff, as given by (5), for the above maturities. In (Demeterfi et al. 1999*b*) the performance of the log payoff as a hedge instrument for the variance swap is studied in the presence of a single down-jump throughout the life of the contract. In equation 42 the authors show that, in this case, the log payoff is worth more than the variance swap. Note that the values of the log payoff in figure 19 dominate the values of the variance swaps, as given by our model. This agrees with (Demeterfi et al. 1999*b*) because our underlying model is among other things an infinite activity jump process with zero intensity for up-jumps. Note also that the effect of jumps forces the price difference between the variance swap and the log payoff to grow with time. This may appear counterintuitive at first because, in the context of the implied volatility smile, one is accustomed to jump effects fading away with time. In this case however the realized variance is cumulative and, with respect to a 5 year variance swap, jumps that occur in 3 years time are equally important to jumps that happen in a week's time. This accounts for the divergent behaviour in figure 19.

As was mentioned in the introduction, the class of HJM-like models for volatility derivatives (see for example (Buehler 2006)) require an entire term structure of variance swaps (like the one in figure 19) to be calibrated. Our approach on the other hand implies one from the vanilla options data (and of course some modelling hypothesis). Note also that joint pdfs for the realized variance and the underlying index are given in figures 13, 14, 15, 16, 17 and 18.

8.5. **The log contract and variance swaps in a market without jumps.** It is well-known that in a market where the underlying follows a continuous stochastic process the fair value of a variance swap is equal to the value of the replicating European option with the logarithmic payoff given in (5) of subsection 2.3 (see for example (Demeterfi et al. 1999*b*) or (Carr & Madan 1998)). In the presence of jumps this equality ceases to hold as can be observed in figure 19. Recall from section 4 that in order to calibrate our model we had to use a non-zero value for the intensities of the down-jumps.

In this subsection the aim is to confirm that the prices of variance swaps and logarithmic payoffs agree in our framework if the randomness of the underlying model is based purely on stochastic and local volatility. To that end we set up a simplified version of the model, with two volatility regimes only, using the following parameters

α	σ_α	β_α	$\bar{\sigma}_\alpha$	ν_α^-	ν_α^+	F_α
0	10.0%	0.70	60%	0	0	100
1	13.5%	0.50	60%	0	0	110

TABLE 3. Parameters for the local volatility regimes in the simplified model without jumps.

$$\mathcal{G}_\alpha^V = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$$

Markov generators for stochastic volatility ($\alpha = 0, 1$).

Note that the jump intensities in this version of the model are deliberately set to zero. The corresponding probability distribution functions for maturities between 6 months and 5 years, as implied by the simplified model, are given in figure 12.

We priced variance swaps, volatility swaps, logarithmic payoffs and structured logarithmic payoffs given by a portfolio of vanilla options in (3) of subsection 2.3, on the notional of one

dollar for maturities between 6 months and 5 years. The results, expressed in volatility terms, can be found in table 4:

Maturity T	$-\frac{2}{T}\mathbb{E}_0\left[\log\left(\frac{S_T}{S_0}\right)\right]$	Portfolio (3)	$\mathbb{E}_0\left[\Sigma_T\right]$	$\mathbb{E}_0\left[\sqrt{\Sigma_T}\right]$
0.5	10.3691	10.3753	10.3699	10.3031
1	10.6368	10.6450	10.6387	10.5327
2	10.9882	11.0003	10.9890	10.8832
3	11.2036	11.2142	11.2013	11.1186
4	11.3494	11.3614	11.3435	11.2602
5	11.4574	11.4691	11.4513	11.3593

TABLE 4. The random variable Σ_T is the annualized realized variance of the underlying process S_T.

It follows from table 4 that the difference between the value of the logarithmic payoff and the fair value of the variance swap, according to our model, is less than 1 volatility point for all maturities. We can also observe the quality of the approximation of the portfolio of options (3) to the log payoff as well as the convexity effect for volatility swaps.

9. Conclusion

In this paper we introduce an approach for pricing derivatives that depend on pure realized variance (such as volatility swaps and variance swaptions) and derivatives that are sensitive to the implied volatility smile (such as forward starting options) within the same framework.

The underlying model is a stochastic volatility model with jumps that has the ability to switch between different CEV regimes and therefore exhibit different characteristics in different market scenarios. The structure of the model allows a straightforward calibration to the implied volatility surface with minimal explicit time-dependence. The stationary nature of the model is best described by the implied forward smile behaviour that can be observed in figures 7, 8, 9 and 10.

The model is then extended in such a way that it captures the realized variance of the underlying process, while retaining complete numerical solubility. Two key ideas that make it possible to keep track of the path information numerically are:

- the observation that path-dependence can be expressed as the lifting of a Markov generator and that
- the lifted process can be chosen so that numerical tractability is retained.

Having obtained the joint probability distribution function for the realized variance and the underlying process, we outline the pricing algorithms for derivatives that are sensitive to the realized variance and the implied volatility within the same model.

References

Albanese, C. & A. Mijatović (2005), 'A stochastic volatility model for risk-reversals in foreign exchange'. Submitted for publication.

Black, F. & M. Scholes (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Bottcher, A. & B. Silbermann (2006), *Analysis of Toeplitz operators*, Springer monographs in mathematics, 2nd edn, Springer.

Breeden, D. & R. Litzenberger (1978), 'Prices of state contingent claims implicit in option prices', *Journal of Business* **51**, 621–651.

Brockhaus, O. & D. Long (1999), 'Volatility swaps made simple', *Risk 2* **1**(1), 92–95.

Buehler, H. (2006), 'Consistent variance curve models'. Working paper, Deutsche Bank.

Carr, P. & D. Madan (1998), Towards a theory of volatility trading, *in* R.Jarrow, ed., 'Volatility: New Estimation Techniques for Pricing Derivatives', Risk publication, Risk, pp. 417–427.

Carr, P., H. Geman, D. B. Madan & M. Yor (2005), 'Pricing options on realized variance', *Finance and Stochastics* **IX**(4), 453–475.

Carr, P. & K. Lewis (February 2004), 'Corridor variance swaps', *Risk* .

Carr, P. & R. Lee (2004), 'Robust hedging of volatility derivatives'. Presentation of Roger Lee at Columbia Financial Engineering seminar.

CBOE, Publication (2003), 'Vix white paper'. see *http://www.cboe.com/micro/vix/vixwhite.pdf*.

Chriss, N. & W. Morokoff (October 1999), 'Market risk for volatility and variance swaps', *Risk* .

Demeterfi, K., E. Derman, M. Kamal & J. Zou (1999*a*), 'A guide to volatility and variance swaps', *Journal of derivatives* **6**(4), 9–32.

Demeterfi, K., E. Derman, M. Kamal & J. Zou (1999*b*), 'More than you ever wanted to know about volatility swaps'. Quantitative Strategies Research Notes, Goldman Sachs.

Detemple, J. & C. Osakwe (2000), 'The valuation of volatility options'. Working paper, Boston University.

Friz, P. & J. Gatheral (2005), 'Valuation of volatility derivatives as an inverse problem', *Quantitative Finance* **5**(6), 531–542.

Grimmett, J. & D. Stirzaker (2001), *Probability and random processes*, 3nd edn, Oxford University Press.

Heston, S. L. & S. Nandi (November 2000), 'Derivatives on volatility: some sample solutions based on observables'. Technical report, Federal Reserve Bank of Atlanta.

Howison, S., A. Rafailidis & H. Rasmussen (2004), 'On the pricing and hedging of volatility derivatives', *Applied Mathematical Finance* **11**, 317–346.

Karatzas, I. & S. E. Shreve (1998), *Brownian motion and stochastic calculus*, Graduate texts in mathematics, 2nd edn, Springer.

Madan, D., P. Carr & E.C. Chang (1998), 'The variance gamma process and option pricing', *European Finance Review* **2**(1), 79–105.

Neuberger, A. (1994), 'The log contract', *The Journal of Portfolio Management* pp. 74–80.

Phillips, R.S. (1952), 'On the generation of semigroups of linear operators', *Pacific Journal of Mathematics* **2**(3), 343–369.

Schoutens, W. (2005), 'Moment swaps', *Quantitative Finance* **5**(6), 525–530.

Windcliff, H., P.A. Forsyth & K.R. Vetzal (2006), 'Pricing methods and hedging strategies for volatility derivatives', *Journal of Banking and Finance* **30**, 409–431.

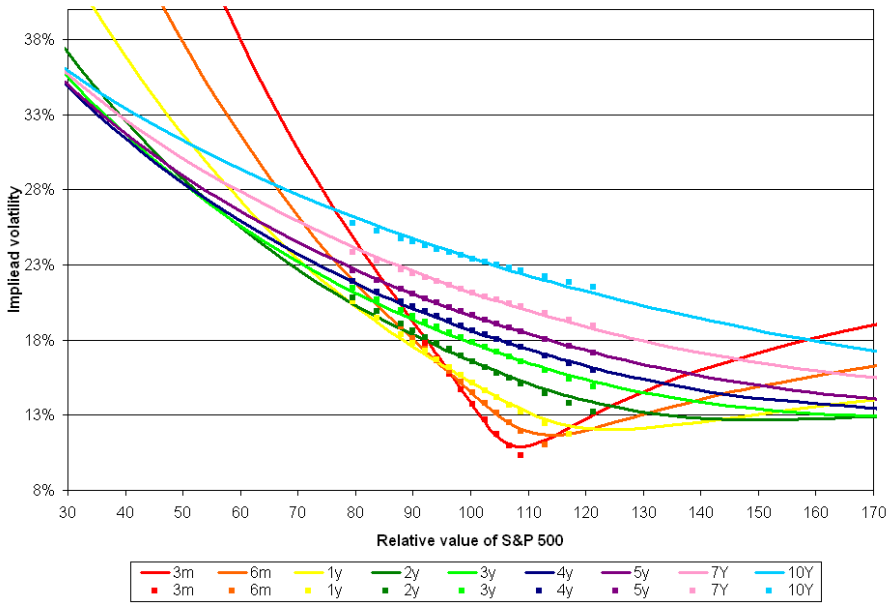

FIGURE 1. Implied volatilities for out-of-the-money European options written on the S&P 500 with maturities between 3 months and 10 years. For each market-specified maturity the squares ■ represent market-implied volatilities. The continuous curves graph the implied volatility of the model as a function of strike. The relative value of S&P 500, with respect to the current level of spot, is plotted along the line of abscisse.

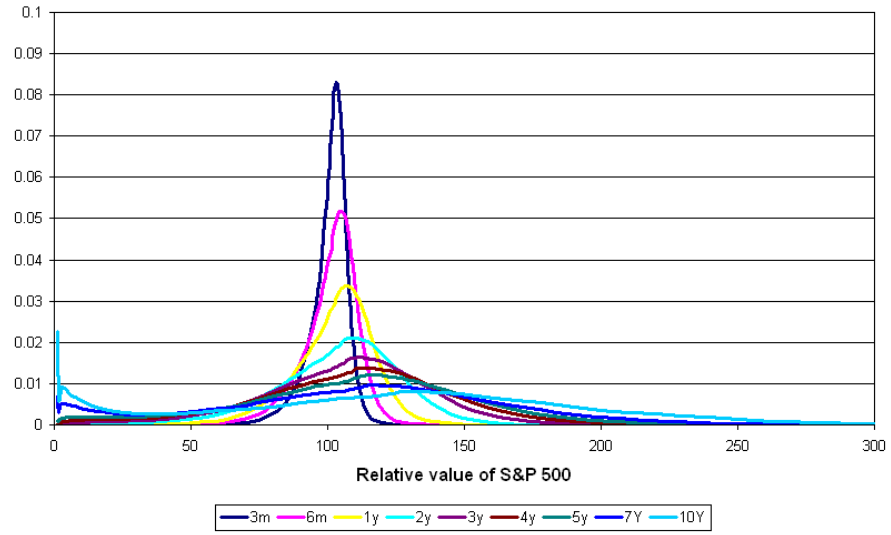

FIGURE 2. Implied probability density function for the S&P 500 under the forward measure. The value of the S&P 500 plotted along the line of abscisse is a relative value (in percent) with respect to the current level of spot. The pdfs between 6 months and 5 years can also be viewed as the rescaled marginals of the joint pdfs in figures 13, 14, 15, 16, 17 and 18.

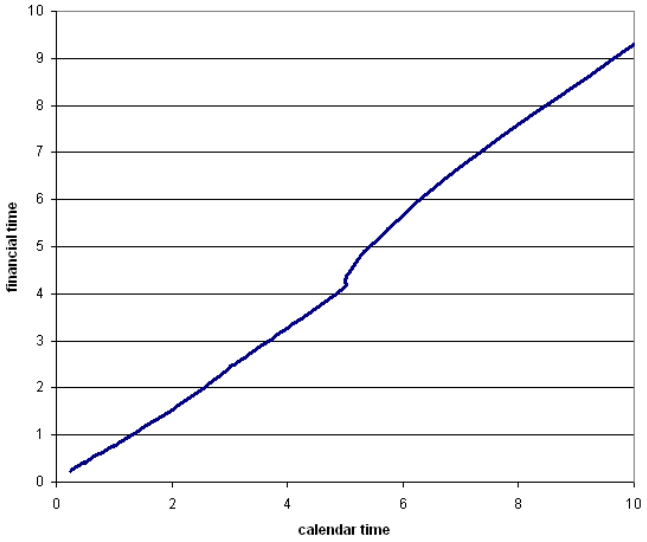

FIGURE 3. Deterministic time-change $f(t)$ (measured in years) as a function of calender time t (also in years). Function f was used in the calibration of the model to the vanilla surface of the S&P 500.

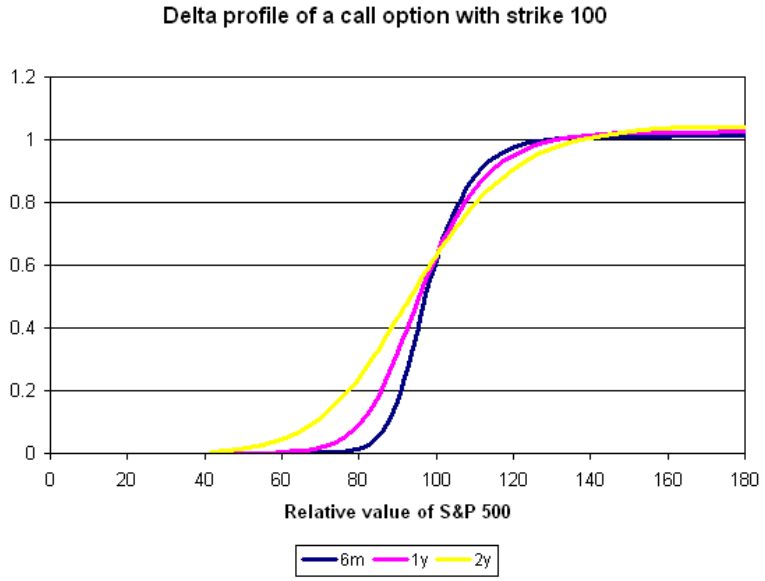

FIGURE 4. Delta profiles of call options on the S&P 500 with maturities between 6 months and 2 years, all struck at 100. We calculate the delta of a call option (i.e. $\Delta(S) = \frac{\partial C}{\partial S}(S)$) for all lattice points S using a symmetric difference as described in subsection 3.5 Notice that we are using the underlying index S and the strike of the options on their relative scales with respect to the level at which the index was trading when the snapshot of the market was taken.

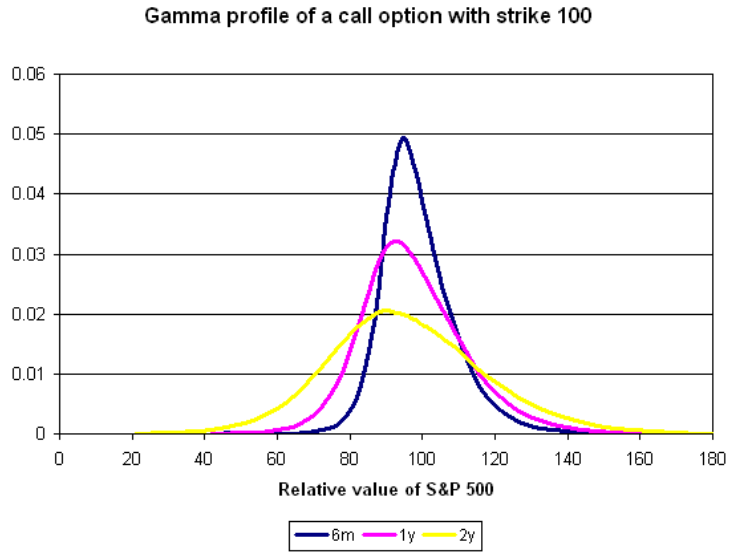

FIGURE 5. Gamma profiles of call options on the S&P 500 with maturities between 6 months and 2 years, all struck at 100. We calculate the gamma of a call option (i.e. $\Gamma(S) = \frac{\partial^2 C}{\partial S^2}(S)$) for all lattice points S using a symmetric difference as described in subsection 3.5. The same comment as in figure 4, about the relative value of the index and the strike, applies.

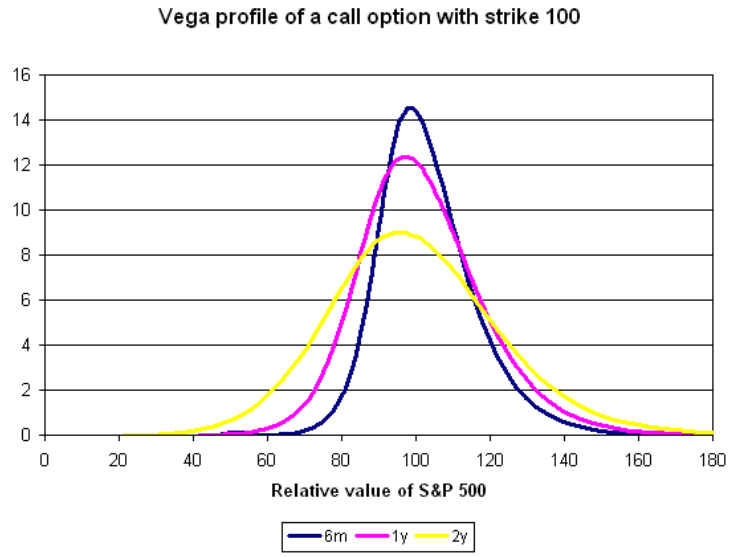

FIGURE 6. Vega profiles of call options on the S&P 500 with maturities between 6 months and 2 years, all struck at 100. We are calculating the vega of a call option by bumping the current volatility regime, repricing the option and plotting the difference from the original option value, for all points on the lattice. Notice that vega and gamma profiles are very similar in shape but different in magnitude, which is consistent with the general market view on the two Greeks.

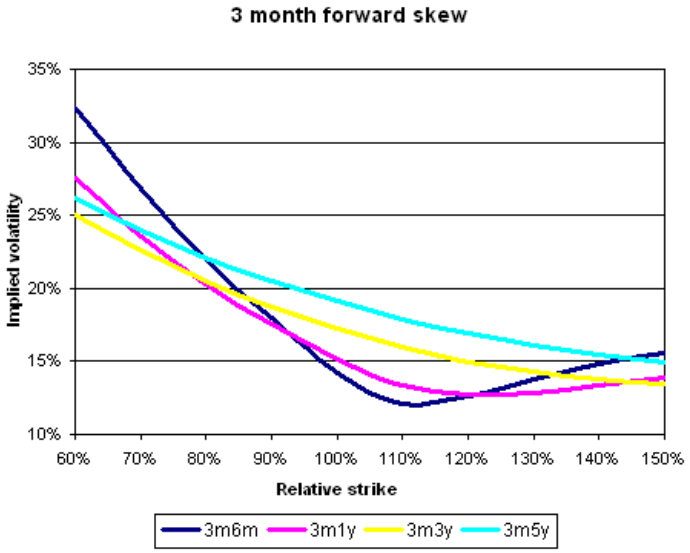

FIGURE 7. Implied forward volatility skews 3 months from now. For example the code 3m1y means that T' equals 3 months and that the expiry T can be calculated as $T = (T' + 1\,\text{year})$, where T', T are as defined in subsection 2.3. Along the line of abscisse we plot the forward strike α (i.e. the "moneyness" of the ordinary call option that the forward-start becomes at time T') as defined in formula (2). The ordinate axis contains the forward volatility values expressed in percentage, as impled by the model.

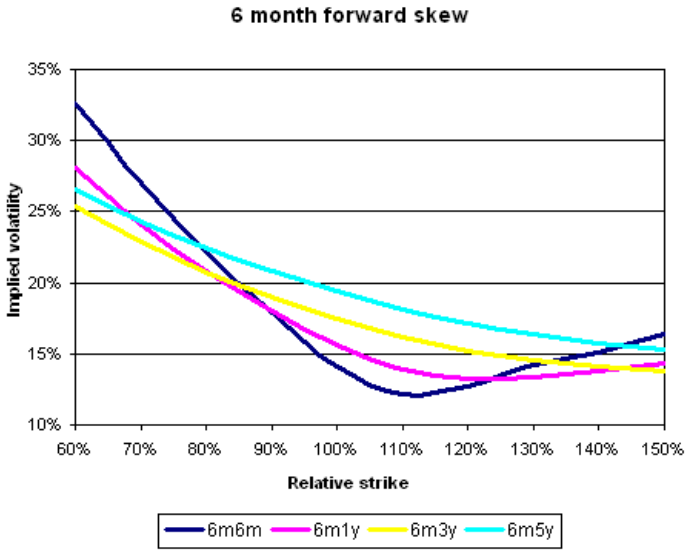

FIGURE 8. Implied forward volatility skews 6 months from now. For example the code 6m1y means that T' equals 6 months and that the expiry T can be calculated as $T = (T' + 1\,\text{year})$, where T', T are as defined in subsection 2.3. Along the line of abscisse we plot the forward strike α (i.e. the "moneyness" of the ordinary call option that the forward-start becomes at time T') as defined in formula (2). The ordinate axis contains the forward volatility values expressed in percentage, as impled by the model.

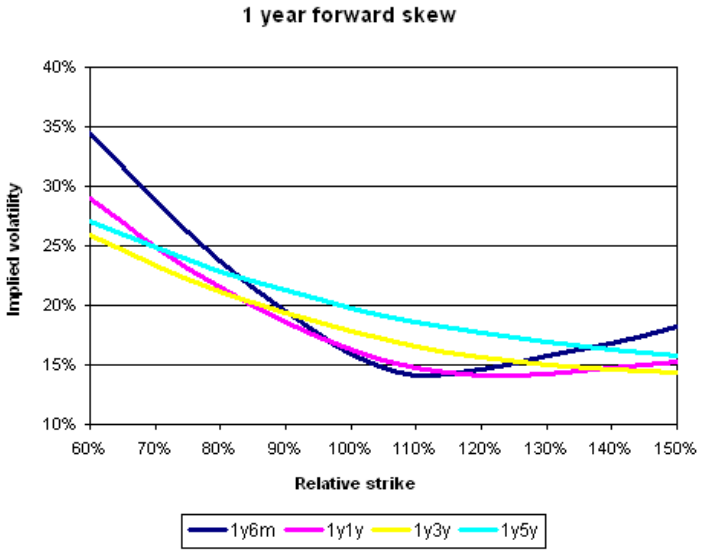

FIGURE 9. Implied forward volatility skews 1 year from now. For example the code 1y6m means that T' equals 1 year and that the expiry T can be calculated as $T = (T' + 6\,\text{months})$, where T', T are as defined in subsection 2.3. Along the line of abscisse we plot the forward strike α (i.e. the "moneyness" of the ordinary call option that the forward-start becomes at time T') as defined in formula (2). The ordinate axis contains the forward volatility values expressed in percentage, as impled by the model.

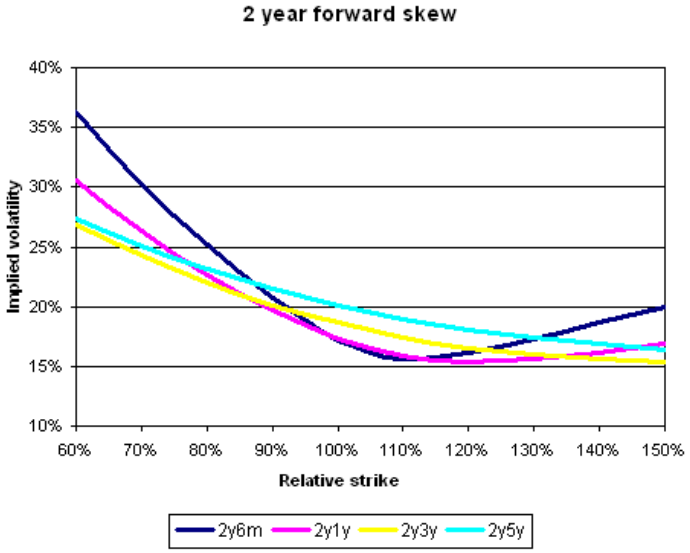

FIGURE 10. Implied forward volatility skews 2 years from now. For example the code 2y6m means that T' equals 2 years and that the expiry T can be calculated as $T = (T' + 6\,\text{months})$, where T', T are as defined in subsection 2.3. Along the line of abscisse we plot the forward strike α (i.e. the "moneyness" of the ordinary call option that the forward-start becomes at time T') as defined in formula (2). The ordinate axis contains the forward volatility values expressed in percentage, as impled by the model.

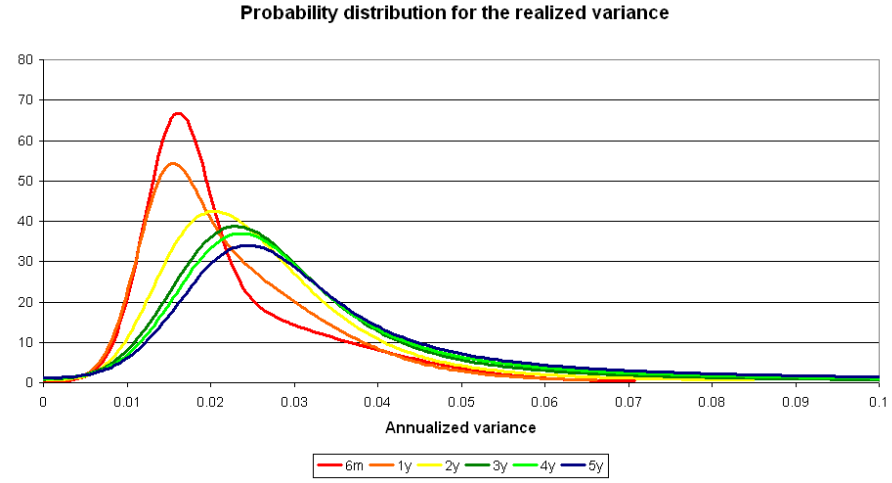

FIGURE 11. Probability distribution functions for the realized variance, quoted in annual terms, for maturities between 6 months and 5 years as given by our model after calibration to the implied volatility surface of the S&P 500 (see section 4 for details). These pdfs are marginal distributions obtained from the joint probability distribution function (32) by integrating it in the dimension of the spot value of the index.

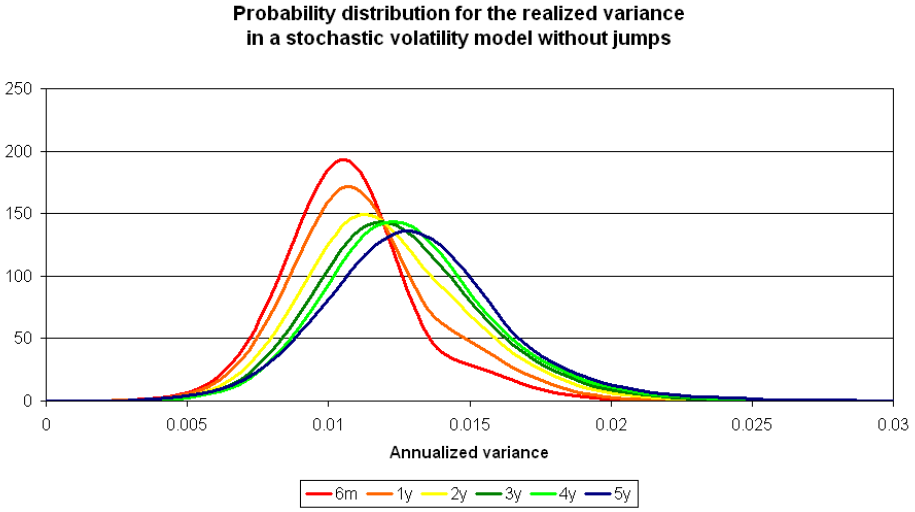

FIGURE 12. Probability distribution functions for the realized variance, quoted in annual terms, for maturities between 6 months and 5 years as given by the model with zero jump intensities, based on two volatility regimes. For the full list of values of the parameters used to specify this model see subsection 8.5.

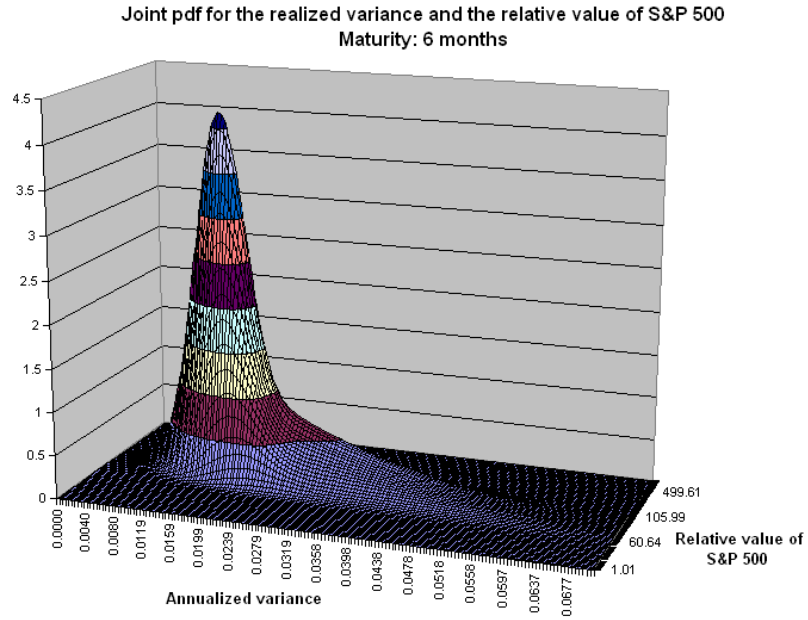

FIGURE 13. Joint probability distribution function for the annualized realized variance and the spot rate of S&P 500 in 6 month's time.

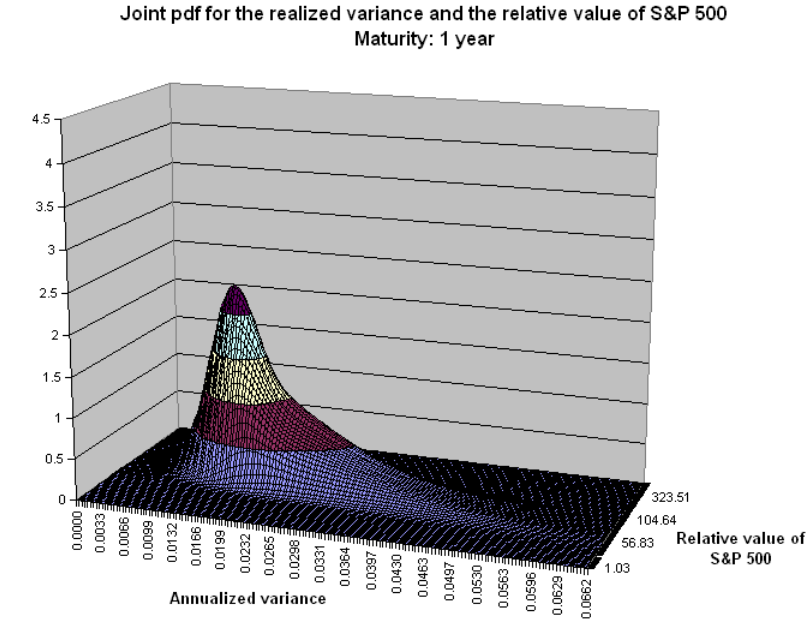

FIGURE 14. Joint probability distribution function for the annualized realized volatility and the spot rate of S&P 500 in 1 year's time.

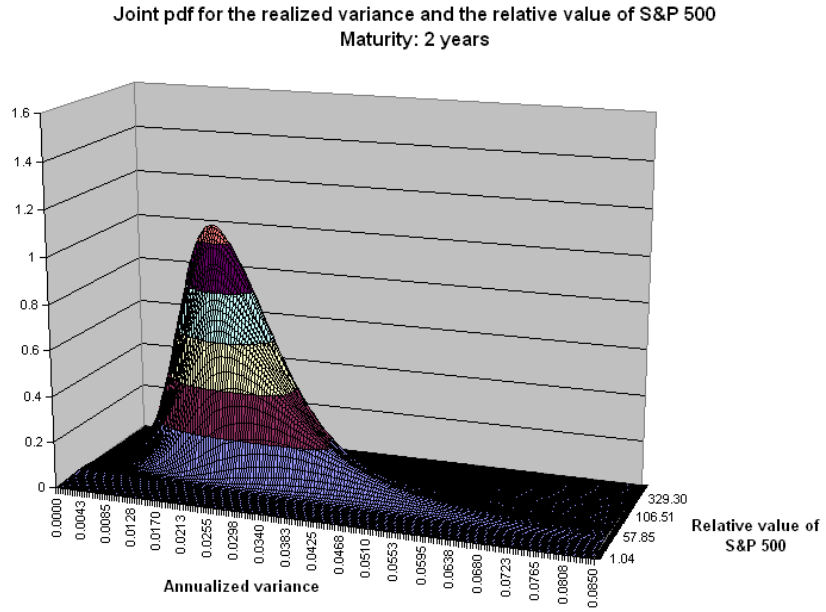

FIGURE 15. Joint probability distribution function for the annualized realized volatility and the spot rate of S&P 500 in 2 years' time.

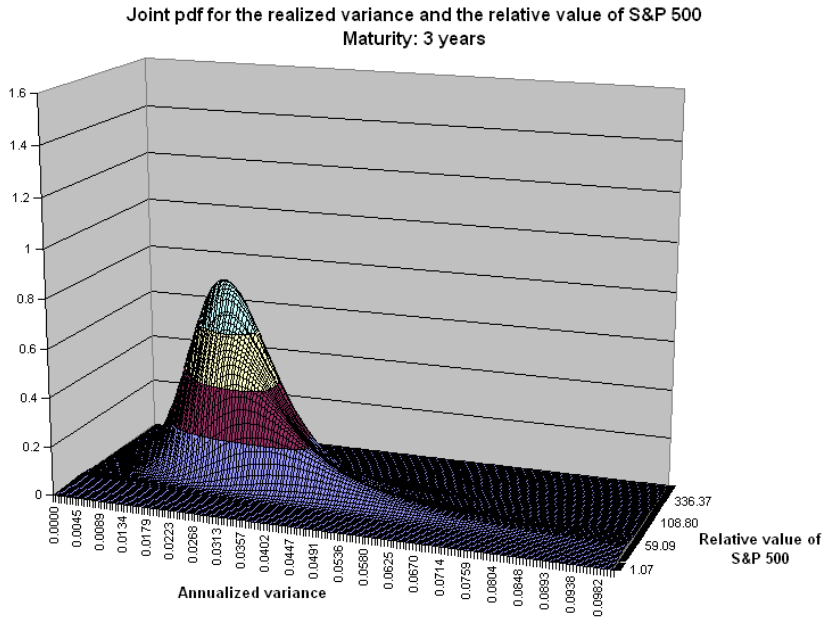

FIGURE 16. Joint probability distribution function for the annualized realized volatility and the spot rate of S&P 500 in 3 years' time.

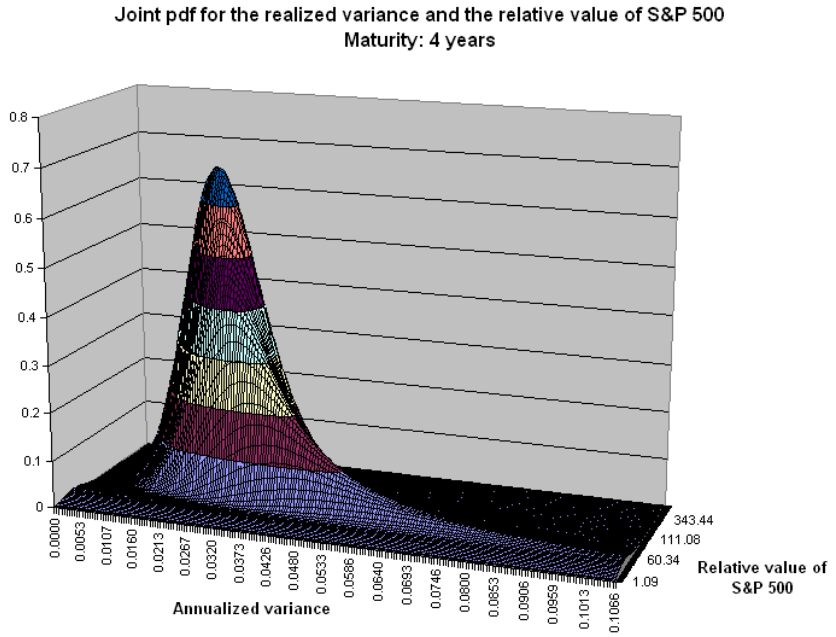

FIGURE 17. Joint probability distribution function for the annualized realized volatility and the spot rate of S&P 500 in 4 years' time.

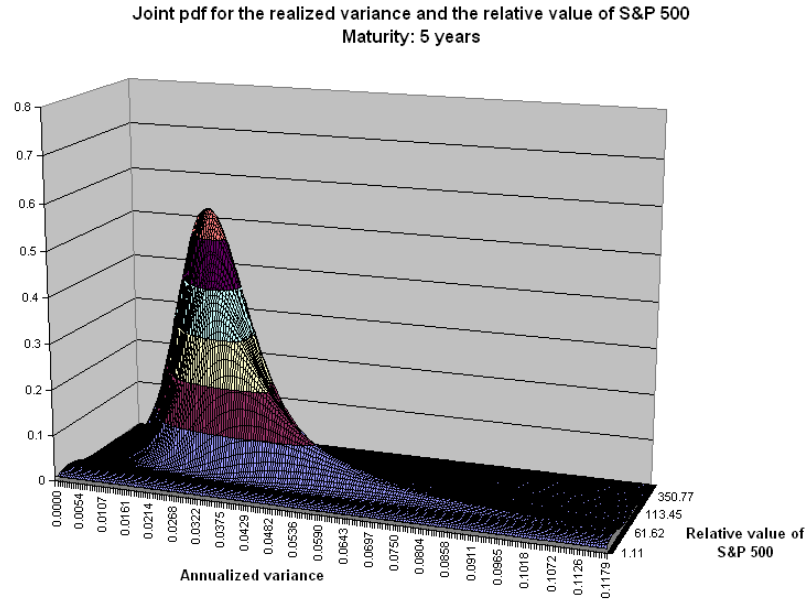

FIGURE 18. Joint probability distribution function for the annualized realized volatility and the spot rate of S&P 500 in 5 years' time.

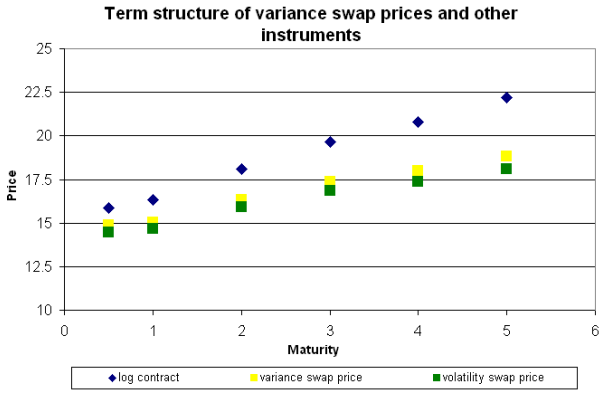

FIGURE 19. Term structures of variance swap prices for maturities between 6 months and 5 years as implied by the vanilla market data using the calibrated model. The delivery prices (i.e. fair strikes) are computed as $\mathbb{E}_0[\Sigma_T]$, where Σ_T is the annualized realized variance for each tenor T. Everything is expressed in terms of volatility, i.e. the prices are in percent and are obtained by taking the square root of the variance. We also plot fair delivery prices for volatility swaps. The convexity bias implied by the model can be clearly observed. The value of the log contract, defined in equation (5), given by the underlying model is also plotted. Observe that this portfolio of options is always worth more than the corresponding variance swap because, in our model, we only allow for down jumps (see section 4). This behaviour is exactly as predicted by the analysis in (Demeterfi et al. 1999*b*) (equation 42) when they added a single down-jump to the underlying process and studied its influence on the static hedge (5) for the variance swap. Note that the increasing difference between the hedging instrument and the variance swap is due to the cumulative nature of realized variance. In other words the current value is effected in the same way by jumps now and jumps just before expiry.

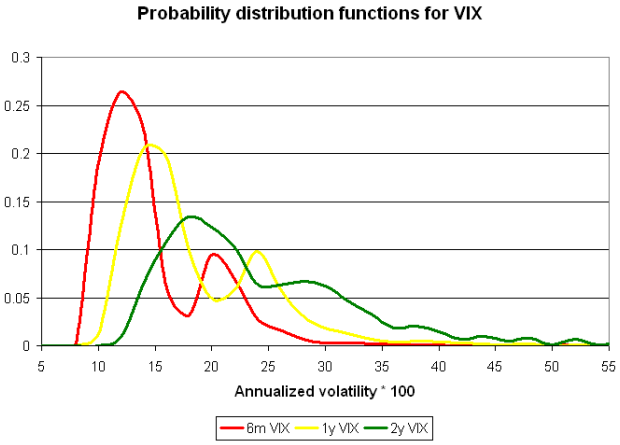

FIGURE 20. Probability distribution functions of the implied volatility index for maturities between 6 months and 2 years. The "irregular" shape of these pdfs is due to the fact that the portfolio of options in formula (3), which define VIX, expires in 1 month after each of the maturities. This implies that the underlying model is unlikely to change the stochastic volatility regime it is in at each maturity (according to generators $\mathcal{G}_\alpha^V$ on page 18, on average a change of regime occurs once every three months). This explains the local maxima which arise close to the values of the constants σ_α in the definition of the local volatility regimes (see table 1 in section 4).

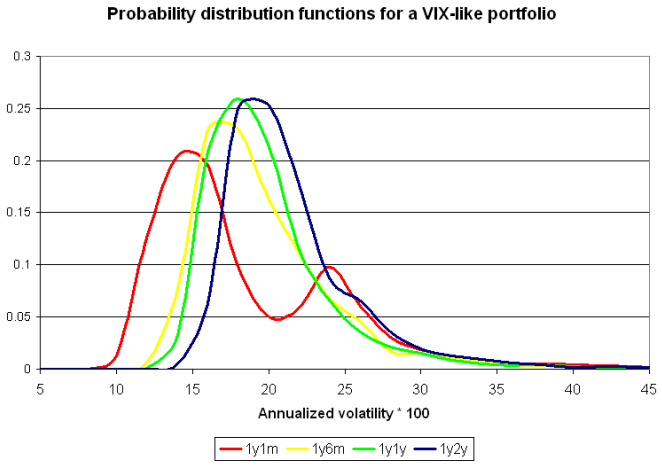

FIGURE 21. Probability distribution functions of a portfolio of forward starting options as in the definition of VIX (see (3) in subsection 2.3), where time t is fixed at 1 year and time T varies form 1 month to 2 years (see subsection 2.3 for the definition of a probability distribution function for a portfolio of forward-starts and the role of parameters t and T). Note that the 1y1m pdf is, according to our definition, the distribution of VIX in 1 year.

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON
E-mail address: claudio.albanese@imperial.ac.uk

SWISS RE CAPITAL MANAGEMENT AND ADVISORY, LONDON, AND IMPERIAL COLLEGE LONDON
E-mail address: harry.lo@imperial.ac.uk

INSTITUTE FOR MATHEMATICAL SCIENCES, IMPERIAL COLLEGE LONDON
E-mail address: a.mijatovic@imperial.ac.uk