

Model Risk in Variance Swap Rates

Carol Alexander and Stamatis Leontsinis

ICMA Centre
Henley Business School, University of Reading

This version: August 4, 2012

Abstract

Different theoretical and numerical methods for calculating the fair-value of a variance swap give rise to systematic biases that are most pronounced during volatile periods. For instance, differences of 10-20 percentage points would have been observed on fair-value index variance swap rates during the banking crisis in 2008, depending on the formula used and its implementation. Our empirical study utilizes more than 16 years of FTSE 100 daily options prices to compare three fair-value variance swap rates. The exchange's variance swap rate formula, used to quote volatility indices such as VIX, has an upward bias induced by Riemann sum numerical integration that empirically outweighs the negative jump and discrete-monitorization biases that are inherent in this fair-value formula. On average, the exchange's methodology provides less accurate predictors of discretely-monitored realised volatility than the approximate swap rate formula introduced in this paper, which we implement using an almost exact analytical integration technique.

JEL Code: G01, G12, G15

Keywords: Model Risk, Variance Swap, Volatility Index, VIX, FTSE 100, VFTSE.

1 Introduction

Variance swaps have been actively traded over the counter since the mid 1990's. Their demand stems from speculation on volatility, equity diversification, dispersion trading and both vega and correlation hedging; see Clark [2010] for more details. Normally it pays to write variance swaps because the variance risk premium is negative, but short variance swap positions are very risky and potential losses can be huge during volatile conditions. Even some of the biggest dealers stopped quoting single-stock variance swaps rates during the banking crisis, because it became impossible to hedge.¹ Volatility swaps limit the dealer's risk better than variance swaps because their pay-offs are linear, not convex, in volatility. But, by the same token, they are not as attractive to purchasers as variance swaps. Moreover, they are more difficult to price and hedge. Hence, soon after the banking crisis in 2008 the variance swap market returned with a growing interest in conditional swaps and swaps with other features that limit the issuer's risk.²

A vanilla variance swap initiated at time t and expiring at $T > t$ is a forward contract which pays a fixed strike \bar{K} and receives the realised variance $\text{RV}_{t,T}$ of the underlying asset returns from time t to time T . The terms and conditions of the contract specify how $\mathcal{V}_{t,T}$ should be calculated at expiry;³ typically it is an average of squared daily log returns on the underlying between times t and T . The variance swap rate (VSR) is the fixed strike \bar{K} ; this is the price at which two counterparties agree to enter the swap.

The pay-off to a buyer of notional N on a vanilla variance swap is⁴ $N [\text{RV}_{t,T} - \bar{K}]$. Hence, the fair value $K_{t,T}$ for the swap strike \bar{K} at initiation is the risk-neutral expectation of $\text{RV}_{t,T}$. Whilst the two parties may enter the swap at any price \bar{K} they can agree upon (and clearly the purchaser will, *ceteris paribus*, seek the lowest price available) each should assess this price

¹The vega exposure of variance swaps increases linearly with the level of volatility, so more options are needed to hedge as volatility increases. However, sufficient vega is often not available in stock options when markets become less liquid, as they can when the stock market crashes.

²For instance in April 2006 JP Morgan introduced corridor variance swaps, which limit the downside risk to both counterparties by paying realised variance only when it is within a pre-defined range; and in March 2009 Société General launched the American variance swap. Since the banking crisis in 2008 it has become standard to cap realised variance at 2.5 times the swap rate.

³For instance, see MorganMarkets [2006].

⁴When banks quote swap rates and realised variances these are typically expressed as annual volatilities. For instance, if an investor buys £75 per basis point notional on a variance swap with six months to maturity at a strike of 23%, and if the realised volatility computed at the swap's expiration six months later turns out to be 25%, the investor receives $\pounds 75 \times 100 \times \frac{25^2 - 23^2}{2} = \pounds 360,000$ from the dealer at expiry.

having some idea of the fair value, $K_{t,T}$. How different could the two parties' assessments of $K_{t,T}$ be?

This paper demonstrates how $K_{t,T}$ may be derived from the current prices of vanilla options using a variety of theoretical and numerical methods. We analyse the theoretical differences and investigate the empirical differences between the various approaches. Model risk arises because the two parties may decide to employ different formulas for $K_{t,T}$, different data filtering methods, and/or different numerical methods for implementing the formula. It is worth noting here that, whilst various theoretical and numerical biases affect the fair value of a discretely monitored variance swap, such biases are irrelevant for the pricing of derivatives on VIX (such as the futures and options that are traded on the CBOE). This is because both the formula used and its numerical implementation methodology are precisely defined in the VIX white paper (available from www.cboe.com/micro/vix/vixwhite.pdf) so the underlying is uniquely defined (and indeed, is quoted on the exchange).

Three possible formulas for computing $K_{t,T}$ are described: the well-known formula based on the first moment of the quadratic variation of the log return which is used by exchanges for quoting volatility indices; a formula introduced by Bakshi et al. [2003] which computes $K_{t,T}$ as the second moment of the log return distribution; and an approximate formula, derived in this paper, where $K_{t,T}$ is obtained from the moment generating function of the price density. We elucidate their differences and their different data-filtering, discrete-monitorization and numerical implementation biases, and provide an empirical assessment of the accuracy of each fair-value VSR as a predictor of realised variance using daily time series for fixed-term FTSE 100 fair-value VSRs of maturities 30, 60, ..., 270 days over a period of more than 16 years, from October 1992 to March 2009.

In the following: Section 2 sets out three different formulas for $K_{t,T}$; Section 3 describes the vanilla option data and its filtering; Section 4 discusses numerical interpolation and integration techniques; Section 5 presents empirical results and Section 6 concludes.

2 Three Fair-Value Formulas

Assume the following diffusion process for the underlying price S_t :

$$\frac{dS_t}{S_t} = rdt + \sigma_t dW_t, \quad (1)$$

where r is the risk-free rate, assumed constant, and W_t is a Brownian motion. The total variance from time t to time T is $\int_t^T \sigma_s^2 ds = \mathcal{V}_{t,T}$ and the continuously monitored realised variance is defined as $\text{RV}_{t,T} = (T - t)^{-1} \mathcal{V}_{t,T}$. Thus, the fair value VSR is

$$K_{t,T} = (T - t)^{-1} \mathbb{E}_t \left[\int_t^T \sigma_s^2 ds \right], \quad (2)$$

where $\mathbb{E}_t[\cdot]$ denotes the expectation at time t under the risk-neutral measure. We now consider three possible representations of $K_{t,T}$ in terms of vanilla call and put option prices, which we label $K_{t,T}^{(1)}$, $K_{t,T}^{(2)}$ and $K_{t,T}^{(3)}$.

Under the geometric Brownian motion (GBM) assumption where σ_t is constant, Neuberger (1994) showed that $\mathcal{V}_{t,T}$ is the expected pay-off to a short position on a contract with pay-off $\log(S_T/S_t)$ at time T . The market is complete so this pay-off, and therefore also the realised variance, can be replicated using vanilla options. Indeed, Demeterfi et al. [1999] prove that:

$$K_{t,T}^{(1)} = 2r + 2(T - t)^{-1} \left\{ e^{r(T-t)} \int_0^\infty \frac{\mathcal{Q}_t(K, T)}{K^2} dK - \frac{F_{t,T} - S_{t,T}^*}{S_{t,T}^*} - \ln \left(\frac{S_{t,T}^*}{S_t} \right) \right\} \quad (3)$$

where $F_{t,T} = e^{r(T-t)} S_t$ is the T -maturity forward price, and $\mathcal{Q}_t(K, T)$ is a fair-value vanilla put or call price of strike K and maturity T . Specifically, for $K < S_{t,T}^*$ it is the price of a vanilla put maturing at time T and for $K > S_{t,T}^*$ it is the corresponding call price. We use the same notation throughout for the separation strike $S_{t,T}^*$, i.e. the boundary separating the out-of-the-money call and put options. Typically this is chosen as the highest strike at or below $F_{t,T}$.

Formula (3) is the first fair-value VSR that we consider. It has been popularised by its widespread adoption amongst exchanges that quote volatility indices and trade futures on these

indices.⁵ We remark that the assumption (1) made by Demeterfi et al. [1999] can be generalised: Carr and Lee [2003] and Jiang and Tian [2005] show that (3) still holds under any type of dynamics for the stochastic volatility process σ_t including a non-zero price-volatility correlation. Carr and Lee [2003] show that adding a Poisson jump to the price process introduces a small bias to the formula (3).⁶ Rompolis and Tzavalis [2009] derive bounds for this bias and demonstrate, via simulations and an empirical study, that price jumps induce a systematic negative bias that is particularly apparent during excessively volatile periods, when (3) significantly underestimates the fair-value swap rate.

This negative jump bias is compounded by another theoretical bias which arises because the typical contract specifies realised variance as a discretely monitored average of squared log returns. As shown by Carr and Lee [2009] the discrete-monitorization bias in (3) is typically negative and, like the jump bias, is most pronounced during volatile periods. However, later on we shall see that these theoretical biases could be outweighed by positive biases induced by numerical errors in the standard implementation of (3) that is employed by exchanges when quoting volatility indices.

An alternative VSR formula computes the second moment of the log return rather than the first moment of the quadratic variation of the log return. Under this second approach, first used by Bakshi et al. [2003], there is no need to assume a specific dynamic process for the underlying asset price, except that it is a semimartingale.⁷ The formula is based on a result of

⁵At the time of writing the VIX, VXN, RVX and VXD are quoted on the CBOE, based on the the S&P500, Nasdaq 100, Russell 2000 and Dow Jones Industrial Average (DJIA) indices respectively. The Euronext-Liffe exchange quotes many volatility indices based on European stock indices, including the VDAX, VSTOXX, VCAC, VSMI and VFTSE. The CBOE defines $S_{t,T}^*$ to be the forward price derived from the put-call parity relationship at the point where the difference between a put and call struck is minimal for each specific maturity. If this happens to be the strike of traded vanilla options, then $Q_t(K, T)$ is taken to be the average of the put and call prices at this strike.

⁶Their results are proved by linking the moment generating function of the quadratic variation of the log return with the expected pay-off to a power option, and hence decomposing it into parts that may be replicated with tradable instruments. Then (3) is the first moment of the quadratic variation, i.e. the expected realised variance. Friz and Gatheral [2005] used Carr and Lee's results to derive an expression for the expectation of the square root of the quadratic variation of the returns, which is useful for pricing volatility derivatives. Gatheral [2006] also arrives at (3) using a model-free decomposition of the pay-off to a power option, by computing the first moment of the quadratic variation of the returns process assuming zero price-volatility correlation.

⁷Typical variance swap contracts define the realised variance to be an average of the squared daily log returns over all trading days since initiation of the variance swap. As such, the variance of the T -period log return should be the sum of the variances of daily log returns, as would be the case under the assumption of independence of daily log returns. Beyond this we may also remark that the focus on log returns rather than price changes is consistent with a geometric rather than an arithmetic process.

Carr and Madan [2001] that any twice differentiable function $f(S)$ written as a function of the price S can be expressed as:

$$f(S) = f(\bar{S}) + (S - \bar{S})f'(\bar{S}) + \int_0^{\bar{S}} f''(K)(K - S)^+ dK + \int_{\bar{S}}^{\infty} f''(K)(S - K)^+ dK. \quad (4)$$

Applying (4) to a contract with payoff $f(S) = r_{t,T}^2$ where $r_{t,T} = \ln(S_T/S_t)$ is the log return from t to T , and using a fourth-order Taylor expansion of the moment generating function of the underlying price at maturity T , Bakshi et al. [2003] derive the following expression for the risk-neutral mean of the log return:

$$\begin{aligned} \mu_{t,T} &= e^{r(T-t)} + \frac{V_{t,T}}{2} - \frac{W_{t,T}}{6} - \frac{X_{t,T}}{24} - 1, \\ V_{t,T} &= 2e^{r(T-t)} \int_0^{\infty} \left(1 + \ln \left[\frac{S_{t,T}^*}{K}\right]\right) \frac{\mathcal{Q}_t(K, T)}{K^2} dK, \\ W_{t,T} &= -3e^{r(T-t)} \int_0^{\infty} \frac{2 \ln \left[\frac{S_{t,T}^*}{K}\right] + \left(\ln \left[\frac{S_{t,T}^*}{K}\right]\right)^2}{K^2} \mathcal{Q}_t(K, T) dK \\ X_{t,T} &= 4e^{r(T-t)} \int_0^{\infty} \frac{3 \left(\ln \left[\frac{S_{t,T}^*}{K}\right]\right)^2 + \left(\ln \left[\frac{S_{t,T}^*}{K}\right]\right)^3}{K^2} \mathcal{Q}_t(K, T) dK. \end{aligned}$$

Here $\mu_{t,T}$ is the risk-neutral expected value of the log return $r_{t,T}$ and $V_{t,T}$ is the risk-neutral expected value of the contract with payoff $r_{t,T}^2$. This yields a second formula for the fair-value VSR, viz.

$$K_{t,T}^{(2)} = (T - t)^{-1} (V_{t,T} - \mu_{t,T}^2). \quad (5)$$

More recently, Rompolis and Tzavalis [2009] have derived the risk-neutral characteristic function of any random pay-off defined on a semimartingale and show that (5) may be derived from this when the pay-off is the log return.

Thirdly, we derive an approximate fair-value VSR based on the risk-neutral measure for S_T , as seen from time t . Using the well-known result of Breeden and Litzenberger [1978], the risk-neutral moment generating function for S_T , denoted $\mathcal{M}_{t,T}(\lambda) = \mathbb{E}_t[\exp(\lambda S_T)]$ may be expressed as

$$\mathcal{M}_{t,T}(\lambda) = \int_{K=0}^{\infty} e^{\lambda K} \frac{\partial^2 \tilde{C}(K, T)}{\partial K^2} dK,$$

where for brevity we write $\tilde{C}(K, T) = e^{r(T-t)}C(K, T)$ and later on similarly for $\tilde{P}(K, T)$, the accounted put price. Integrating by parts twice and using the following properties of call option prices:

$$\lim_{K \rightarrow 0} \tilde{C}(K) = S_T, \quad \lim_{K \rightarrow +\infty} \tilde{C}(K) = 0, \quad \lim_{K \rightarrow 0} \frac{\partial \tilde{C}(K)}{\partial K} = -1 \quad \text{and} \quad \lim_{K \rightarrow +\infty} \frac{\partial \tilde{C}(K)}{\partial K} = 0,$$

we obtain

$$\begin{aligned} \mathcal{M}_{t,T}(\lambda) &= e^{\lambda K} \frac{\partial \tilde{C}(K, T)}{\partial K} \Big|_0^\infty - \lambda \int_{K=0}^\infty e^{\lambda K} \frac{\partial \tilde{C}(K, T)}{\partial K} dK \\ &= 1 - \lambda e^{\lambda K} \tilde{C}(K, T) \Big|_0^\infty + \lambda^2 \int_{K=0}^\infty e^{\lambda K} \tilde{C}(K, T) dK. \end{aligned}$$

Thus, the moment generating function for S_T is:

$$\mathcal{M}_{t,T}(\lambda) = 1 + \lambda F_{t,T} + \lambda^2 \int_{K=0}^\infty e^{\lambda K} \tilde{C}(K, T) dK. \quad (6)$$

The n^{th} moment of the distribution of S_T is obtained by taking the n^{th} derivative of (6) with respect to λ and setting $\lambda = 0$. Thus, for $n = 1, 2, \dots$

$$\begin{aligned} \mathbb{E}_t[S_T^n] &= \frac{d^n \mathcal{M}_{t,T}(\lambda)}{d\lambda^n} \Big|_{\lambda=0} \\ &= \left[\begin{aligned} &1_{\{n=1\}} F_{t,T} + n(n-1) \int_{K=0}^\infty e^{\lambda K} K^{n-2} \tilde{C}(K, T) dK + \\ &2n \lambda \int_{K=0}^\infty e^{\lambda K} K^{n-1} \tilde{C}(K, T) dK + \lambda^2 \int_{K=0}^\infty e^{\lambda K} K^n \tilde{C}(K, T) dK \end{aligned} \right] \Big|_{\lambda=0} \\ &= 1_{\{n=1\}} F_{t,T} + n(n-1) \int_{K=0}^\infty K^{n-2} e^{\lambda K} \tilde{C}(K, T) dK. \end{aligned} \quad (7)$$

Using the put-call parity relationship $\tilde{C}(K, T) - \tilde{P}(K, T) = F_{t,T} - K$ we transform (6) and (7) to include only the out-of-the-money put and call options, with separation strike $S_{t,T}^*$. After

some straightforward calculations (6) may be written:

$$\mathcal{M}_{t,T}(\lambda) = e^{\lambda S_{t,T}^*} (1 + \lambda F_{t,T} - \lambda K) + \lambda^2 e^{r(T-t)} \int_{K=0}^{\infty} e^{\lambda K} \mathcal{Q}_t(K, T) dK.$$

Differentiating twice w.r.t. λ and setting $\lambda = 0$ gives the second moment:

$$\mathbf{E}_t[S_T^2] = 2(F_{t,T} - S_{t,T}^*) S_{t,T}^* + (S_{t,T}^*)^2 + 2e^{r(T-t)} \int_{K=0}^{\infty} \mathcal{Q}_t(K, T) dK$$

Since $\mathbf{E}_t[S_T] = F_{t,T}$, the second central moment of the risk-neutral density for S_T is:

$$\mathbf{V}_t[S_T] = - (F_{t,T} - S_{t,T}^*)^2 + 2e^{r(T-t)} \int_{K=0}^{\infty} \mathcal{Q}_t(K, T) dK.$$

In the case of a deterministic volatility geometric diffusion with Brownian W_t we have:

$$\mathbf{E}_t[S_T^2] = S_t^2 \exp\left(2r(T-t) - \int_t^T \sigma_s^2 ds\right) \exp\left(2 \int_t^T \sigma_s^2 ds\right) = S_t^2 \exp\left(2r(T-t) + \int_t^T \sigma_s^2 ds\right),$$

because $\mathbf{E}_t[\exp(2\sigma W_T)] = \exp\left(2 \int_t^T \sigma_s^2 ds\right)$. Thus, $\mathbf{V}_t[S_T] = F_{t,T}^2 (\exp[\mathcal{V}_{t,T}] - 1)$, and so setting

$$K_{t,T}^{(3)} = (T-t)^{-1} \left(\frac{2}{F_{t,T}^2} \int_{K=0}^{\infty} \mathcal{Q}_t(K, T) dK - \left(\frac{F_{t,T} - S_{t,T}^*}{F_{t,T}} \right)^2 \right) \quad (8)$$

yields a VSR that is based on a fair-value expression for $\exp[\mathcal{V}_{t,T}] - 1$, not a risk-neutral expectation of $\mathcal{V}_{t,T}$. They are only the same to first order approximation.

Note that (8) has a computational advantage because the integration is over option prices alone. By contrast, the division by the square of the option strike within the integrals in both (3) and (5) places a greater emphasis on the prices of options at the lowest strikes in the integration range. Since these deep out-of-the-money puts are not heavily traded their final trades are likely to occur well before the market closes. Hence, formulas (3) and (5) will be more sensitive than (8) to the mispricing of low strike options.

In practice the realised variance is not continuously monitored, it is monitored discretely. That is, the terms and conditions of the majority of variance swaps states that:

$$\text{RV}_{t,T} = T_d^{-1} V_{t,T}^d = T_d^{-1} \sum_{i=t}^{T-1} \left(\log \left[\frac{S_{i+1}}{S_i} \right] \right)^2, \quad (9)$$

where T_d is the number of trading days between t and T . We distinguish the discretely monitored total variance $V_{t,T}^d$ from the continuously monitored variance $\mathcal{V}_{t,T}$, and from henceforth we simply denote them V^d and \mathcal{V} , omitting the time subscripts for brevity.

Carr and Lee [2009] proved that $\mathcal{V} = V^d + \zeta$ where the dominant term in the expansion of ζ is proportional to $\sum_{i=t}^{T-1} \left[\frac{S_{i+1} - S_i}{S_i} \right]^3$. They note that when S is an equity or index price ζ is typically large and negative during volatile periods, and smaller and possibly positive during stable trending periods. Thus we can set $\zeta = \mathcal{V}x$, so that x is the bias expressed as a proportion of the continuously monitored variance, and x is typically negative during volatile periods but can be positive during tranquil periods. Similarly (5) also under-prices variance swaps when the risk-neutral expectation of the sum of the cubed daily returns is negative. But what can be said about the discrete-monitorization bias for (8)?

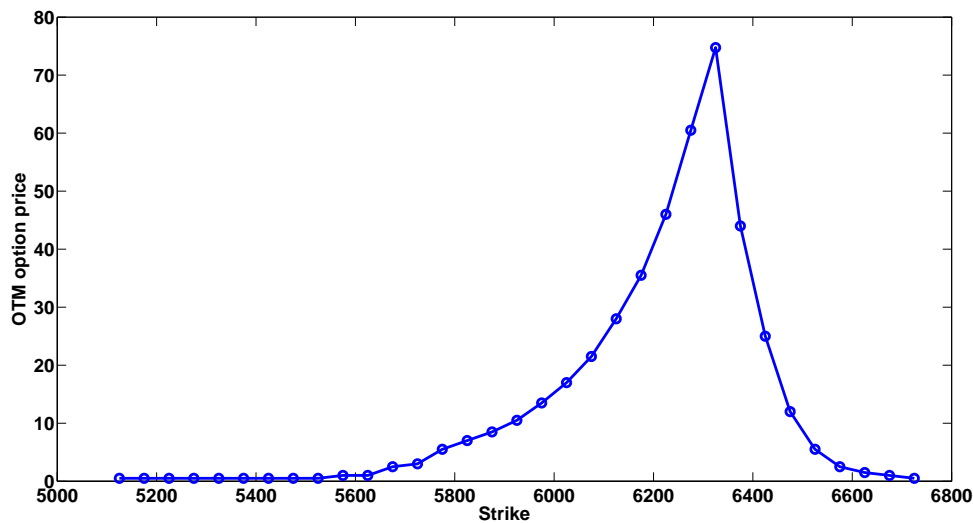
To investigate the discrete-monitorization bias for (8) set $e^{\mathcal{V}} - 1 = V^d + \mathcal{V}y$, where the proportional bias y for (8) may be written $y = x + [\mathcal{V}/2 + \mathcal{V}^2/3! + \mathcal{V}^3/4! + \dots]$, on substituting $V^d = \mathcal{V} - \mathcal{V}x$. The term in brackets is positive and increasing with realised variance, so during volatile periods when x is typically negative, y is smaller and could even be positive; and during stable periods if x is positive, y will be larger and positive.

Finally, we remark that one can always consider a density for S_T but it only makes sense to consider the density of $\log(S_T/S_0)$ when the variance swap is on an index or a tradable asset. If S were an interest rate then realised variance should be defined on daily changes, corresponding to an arithmetic process; interest rates are already a return, and they can go to zero, so the concept of log (or ordinary) returns on interest rates makes no sense.

3 Data

We aim to provide an empirical comparison of the three alternative formulae (3), (5) and (8) for the fair value of a variance swap. To this end we shall employ NYSE Euronext data on closing prices for all vanilla options traded on the FTSE 100 index from 2 October 1992 until 17 March 2009. The total number of option prices is 2,352,916 and, although there are roughly equal numbers of in-the-money (ITM) and out-of-the-money (OTM) option prices, the trading volume on OTM options was approximately seven times greater than the volume traded in ITM options, so we use only OTM calls and puts for the analysis.

Figure 1: OTM Option Prices as a Function of Strike



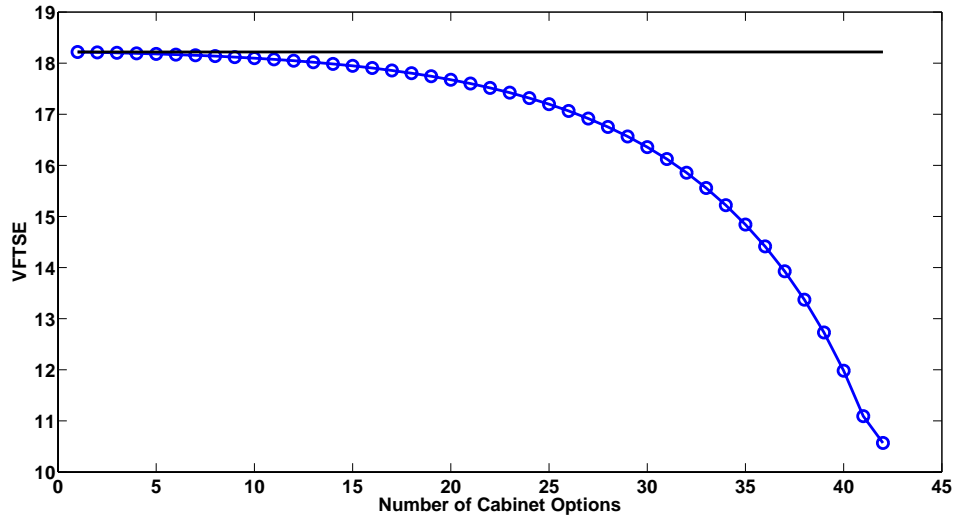
Note: 30-day option prices are created by Hermite spline interpolation on implied variances between the April 2007 and May 2007 expiry dates, as explained in Section 4.

Figure 1 depicts the distribution of FTSE 100 vanilla option prices by strike, with 30 days to maturity on 30 March 2007. The notable negative skew is only a characteristic during tranquil periods – during turbulent positively skewed functions are more common.

‘Cabinet’ options are those options with the minimum price that is specified by the exchange (in this case, £0.5). The name derives from the days when such paper was locked in a cabinet (cupboard) as the trading on these options is virtually nil. Typically, such options are deep OTM put options, the underlying price having risen considerably since the options were first quoted. The database contains 47,472 cabinet option prices, i.e. approximately 2% of the data set. The great majority of these have maturity less than 30 days and moneyness less than 0.9.

Their trading volume is very low and their model price is likely to be less than the exchange’s minimum price, but whilst the open interest is positive the exchange is obliged to quote at least the minimum price.

Figure 2: Effect of Cabinet Options on the VFTSE Evaluation



Note: This figure depicts the effect that the number of cabinet options have at the valuation of the VFTSE by imitating the long left tail of function $\mathcal{Q}(K, T)$. We select the OTM options on 30 March 2007 and then progressively substitute the value of the deepest OTM put option with the value of £0.5, which is the lowest possible tradable value for an option on the FTSE 100 index. The black line is the true value of the VFTSE 30 and the blue line is the VFTSE 30 as a result of this process.

Thus, including these (mostly deep OTM puts) cabinet options in the analysis would induce a negative bias on the VSR. To gain some idea of the size of this bias, Figure 2 depicts the value of the VSR on 30 March 2007, computed using the CBOE methodology, which is fully discussed in the next section. We call this rate the ‘VFTSE’ because formula (3) has been adopted by Euronext for quoting the VFTSE 30-day volatility index. The black line is the value of the VFTSE obtained without cabinet options and the blue line depicts the VFTSE as we progressively set the value of the deepest OTM put option to £0.5. With up to 10 cabinet option prices the bias remains negligible. It is highly unlikely that there would be more than 10 cabinet option prices on any day; nevertheless, we removed these option prices from our database before commencing the empirical work.

4 Numerical Methods

The VSR formulae include two integrals containing an infinite number of OTM put and call options on a continuum of strikes, which has to be approximated to include only the discrete strikes of traded options that are available. For the empirical construction of volatility indices exchanges employ the following discrete approximation to formula (3).

$$K_{t,T} \cong \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} \mathcal{Q}_i(K_i, T) - \frac{1}{T} \left(\frac{F_{t,T}}{K_0} - 1 \right)^2, \quad (10)$$

where $\mathcal{Q}(K_i, T)$ is the price of the i^{th} OTM option with strike K_i and maturity T , and K_0 is the first strike available at or below $F_{t,T}$. Here $\Delta K = (K_{i+1} - K_{i-1})/2$ for strikes straddling the i^{th} OTM option, $\Delta K = K_2 - K_1$ for the lowest strike and $\Delta K = K_n - K_{n-1}$ for the highest strike.

Jiang and Tian [2005, 2007] investigate the truncation, discretization and numerical integration errors in (10). The truncation error is due to the limited strike range whilst the discretization error, which is linked to the numerical integration error, stems from having only a discrete set of strikes available. The truncation error induces a downward bias that is negligible except when a volatility spike is not instantaneously matched by an increase in option price range, and it is typically dominated by an upward bias induced by the strike discretization and the numerical integration technique.

The summation of $\mathcal{Q}(K, T)$ in (10) is a lower Riemann integral which is downward (upward) biased when the function is monotonically increasing (decreasing). So the integral of $\mathcal{Q}(K, T)$ in (3) is under-estimated for the OTM puts and over-estimated for the OTM calls. A negative skew in $\mathcal{Q}(K, T)$ leads to an over-estimation of the integral over calls that outweighs the underestimation of the integral over puts. This creates a net upward bias that becomes more pronounced as the negative skew increases, and as the interval between strikes increases. Jiang and Tian [2007] also demonstrate that an additional upward bias induced by the deviation of the separation strike from the forward price of the index can lead to a significant over-estimation of the fair-value VSR.

Typically, the net upward bias induced by the numerical evaluation of the formula will be

counter-balanced by a negative discrete-monitorization bias during volatile periods, when both biases are greatest. The net result could be a VSR that is actually a fairly accurate predictor of discretely monitored realised variance. We shall investigate whether this could be the case in the next section.

Jiang and Tian [2007] also explore an alternative numerical integration technique based on approximating the integrand using cubic splines, and they use a numerical procedure to demonstrate that it is far more accurate than a simple Riemann sum approximation. A spline is defined as a set of piecewise polynomials passing through n ‘knot points’ which satisfy certain criteria. An exhaustive list of possible spline functions is surveyed by Fritsch and Carlson [1980] and by De Boor [2001]. The most common restriction for cubic splines is that the second derivative of each polynomial is equal to that of its neighbouring polynomial at the knot points. Typically, the second derivative of the first and the last polynomial is set to 0, thus requiring the solution of a tri-diagonal system of $n - 2$ equations. One of their most appealing properties is that they produce zero fitting error when the knots are at the observed points. They are suitable for interpolation of options prices over strikes because they are always twice differentiable, whereas some other splines might not be.

Considering (8) for each fixed T , and henceforth dropping notation of dependence on T for brevity, the function $Q(K)$ is approximated by a cubic polynomial over the interval $[K_i, K_{i+1}]$. If there are n available strikes at a date t for the given maturity T , $Q(K)$ is divided into $n - 1$ third order polynomials of the form:

$$q_i(K) = a_i + b_i(K - K_i) + c_i(K - K_i)^2 + d_i(K - K_i)^3,$$

for $i = 1 \dots n - 1$. On specifying the standard continuity conditions for $Q(K, T)$ the coefficients a_i , b_i , c_i and d_i are obtained by solving a system of $n - 2$ recurrence equations in $n - 2$ unknowns. Even when insufficient strikes are available for knot points and so these strikes are interpolated, a numerical cubic spline integration approach will produce accurate results, as demonstrated by Jiang and Tian [2007].

For the case that $\mathcal{Q}(K)$ itself is the integrand, i.e. when implementing the VSR (8):

$$\int_{K=K_i}^{K_{i+1}} \mathcal{Q}(K) dK = \sum_{i=1}^{n-1} \int_{K=K_i}^{K_{i+1}} q_i(K) dK. \quad (11)$$

Here numerical integration is not necessary, as each integral on the right hand side can be solved analytically, giving the following value for the total integral in (11):

$$\sum_{i=1}^{n-1} \frac{1}{12} (K_{i+1} - K_i) \{12a_i + (K_{i+1} - K_i) [6b_i + (K_{i+1} - K_i) (4c_i + 3d_i (K_{i+1} - K_i))]\}. \quad (12)$$

Given the liquidity of FTSE 100 options and the plethora of strikes available we are able to evaluate integrals of $\mathcal{Q}(K, T)$ almost exactly by applying above to the available strikes at each maturity T .

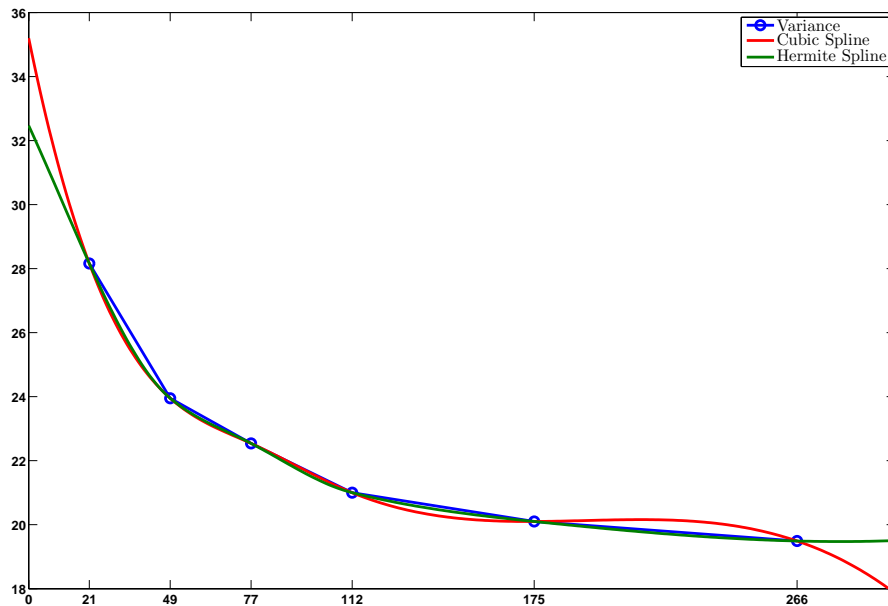
The vanilla options in our data set have fixed maturity dates, so interpolation and extrapolation of these prices to different maturities is required for computing VSRs with fixed terms (in our case, we choose 30, 60, . . . , 270 days). This can be performed by interpolating/extrapolating the vanilla option prices themselves, by interpolating/extrapolating their implied volatilities or variances, or by computing the VSRs for all maturities of the given options and then interpolating/extrapolating these.⁸ Exchanges employ the following simple method: OTM put and call options are allocated to maturity baskets according to their time to expiration; then sub-indices for each expiration date available on that day are calculated, using formula (10); then one obtains the desired maturities for the term structure of VSRs by linearly interpolating and extrapolating the variance term structure implied by these indices; finally we take the square root of the interpolated variance and multiply by 100 to quote the index in percentage points.

Linear interpolation across implied variances ensures no calendar arbitrage, but empirically, implied variance does not grow linearly with term except at long maturities. Instead we apply Hermite cubic spline interpolation across the variances implied by the VSR sub-indices, where each segment is interpolated using a cubic Hermite polynomial of the form

⁸Of course, the computation of fixed maturity VSRs typically also requires interpolation over discount rates, but which type of interpolation is applied makes relatively little difference to the result. Empirical results are available from the authors on request.

$H_3(x) = -e^{x^2/2} \frac{d^3}{dx^3} e^{-x^2/2}$. By construction Hermite splines are smoother than cubic splines because the slopes at the knots is a harmonic mean of the slopes of the polynomials neighbouring each knot, or zero at the ends. Thus Hermite splines have a ‘shape preserving’ feature not shared by cubic splines, which prevents the interpolant from shooting up or down in response to small pricing errors in vanilla options. As such, they are more suitable than cubic splines for extrapolation.

Figure 3: Spline vs Hermite Fitted Implied Volatility Term Structures on 30 March 2007



To illustrate this, Figure 3 compares the two fits to implied variances depicting their annualized square roots, i.e. the implied volatilities. Clearly, cubic splines are too flexible; moreover, the choice of spline matters, as quite different volatility term structure shapes can result. Hermite cubic spline interpolation and extrapolation results must, of course, be checked for no calendar arbitrage. We found that this condition was violated less than 0.2% of days in our sample, i.e. on only 9 out of more than 400,000 days. In this case we simply replaced the Hermite spline with linearly interpolated variances.

5 Empirical Results

Table 1 compares the value of $\int Q(K, T)dK$ that is approximated using the lower Riemann sum (as used by exchanges) and cubic splines, with the analytic solution (12) on 30 March 2007 and 16 October 2008. These two dates were chosen as representative of relatively calm and volatile periods respectively (the FTSE 100 dropped by 5.5% on 16 October 2008). In both cases the lower Riemann sum underestimates the integral over OTM puts, but overestimates the area under OTM calls, because of the negative skew in the risk neutral price density. On 30 March 2007 the net upward bias on the Riemann sum is only $60.4/22,140$, i.e. 0.27%, but on 16 October 2008 (near the height of the banking crisis) the upward bias is greater, at $4,046/409,216 = 0.9\%$.

Table 1: Integral Approximation

	OTM puts	OTM calls	Total
30 Mar 2007			
Lower Riemann Sum	13,825.0	8,337.5	22,162.5
Cubic Spline	15,681.3	6,420.8	22,102.1
16 Oct 2008			
Lower Riemann Sum	224,112.5	189,150.0	413,262.5
Cubic Spline	247,295.4	161,920.8	409,216.2

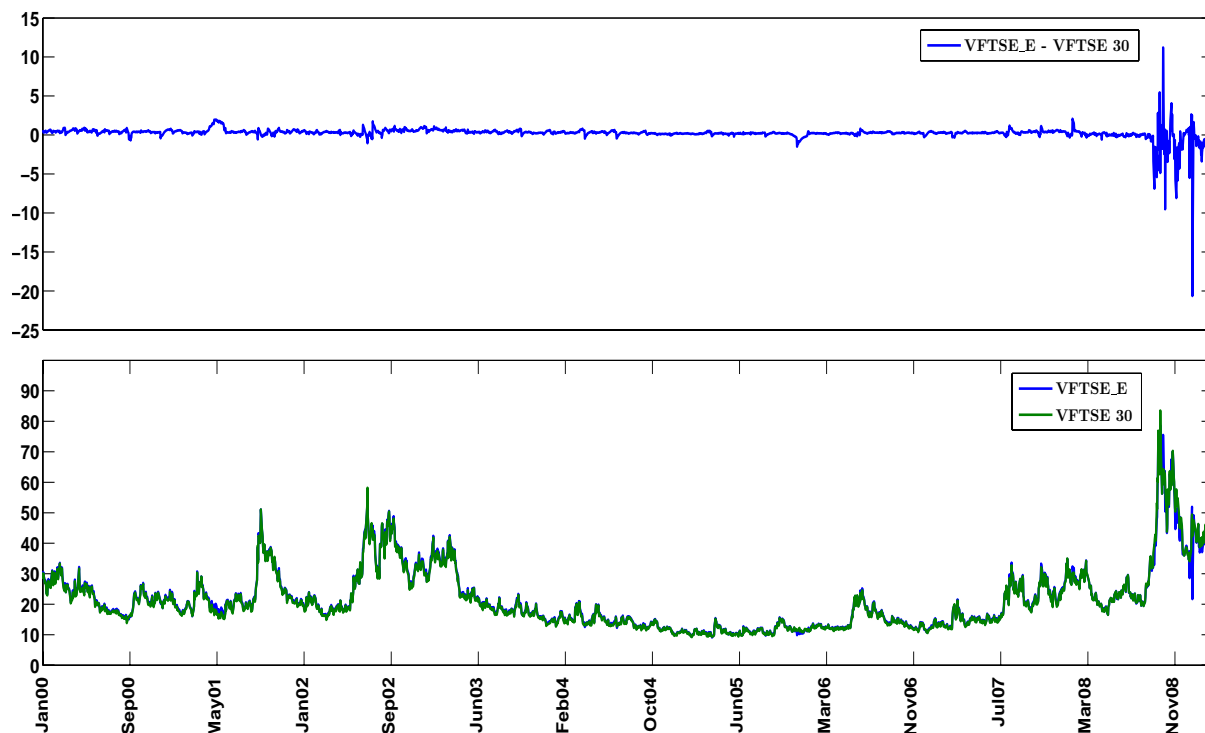
Note: This table evaluates the integral $Q(K, T)$ using two different numerical approaches. The first is the lower Riemann sum and the second uses our closed-form integration after fitting the function using piecewise cubic splines. The result is an upward bias, due to the asymmetry (negative skewness) in the option price function. The dates of the calculations are the 30th March 2007 and the 16th October 2008, randomly chosen as representative dates for low and high volatility, respectively.

The translation of this bias on the integral into VSRs can be very significant.⁹ Figure 4 shows time series of 30-day VSRs based on (10), evaluated using (a) the Riemann sum (denoted VFTSE_E, since this is the index quoted on the Euronext exchange) and (b) cubic spline integration, denoted VFTSE_30. The lower graph depicts the two VSRs and the upper graph depicts their difference. The bias was small until mid 2008, but positive or negative errors of 10 to 20 percentage points occurred on many days during the banking crisis.

We now construct daily time series for VSR term structures with maturities 30, 60, . . . , 270 calendar days, each quoted in annual terms. Figure 5 depicts the daily evolution the VFTSE

⁹In our results all VSRs and realised volatilities are quoted in annual terms, following the market convention.

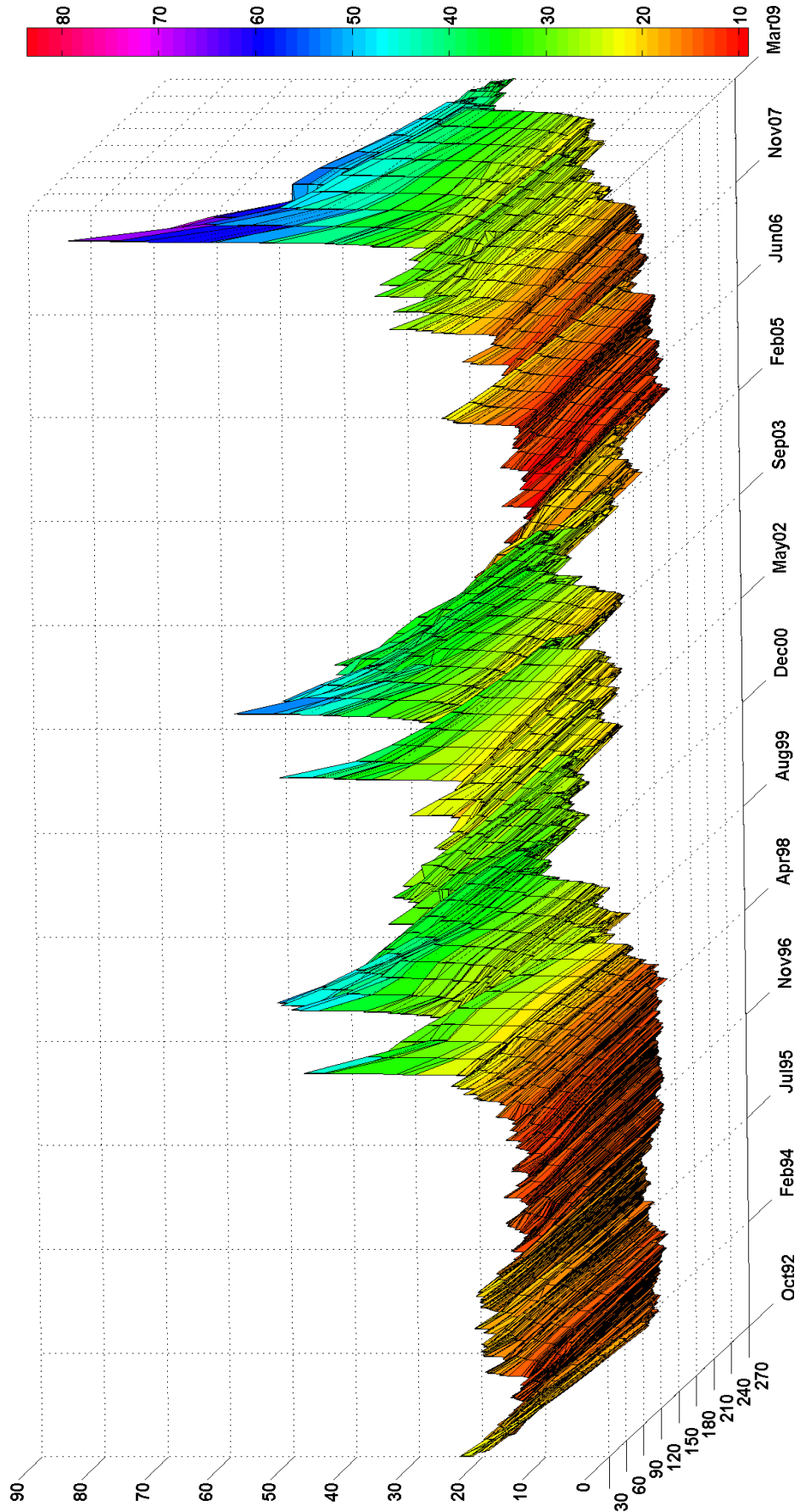
Figure 4: Riemann Sum vs Cubic Spline 30-day VFTSE



Note: This figure presents the pricing errors induced by lower Riemann sum integration of option prices.

term structure between October 1992 and March 2009, based on the exchange's methodology. The variability in the term structure decreases with maturity, as expected, with the short term indices being more responsive to shocks than the longer term indices. During times of low volatility the term structure is in backwardation with short-term volatility being less than longer term volatility. The opposite is the case during volatile periods, when there is an extreme contango. Skewness and kurtosis statistics (not reported, but available on request) indicate that the distribution of these VSRs tends to normality as we increase in term, with the VFTSE 270 being the least skewed and the least leptokurtic. The most pronounced spikes in volatility occurred during the following periods: in July 1997 the East Asian financial crisis; in July 1998 the Russian crisis and LTCM collapse; in September 2001 the World Trade Centre attack; in mid 2002 the burst of technology bubble; in March 2003 the second Gulf War; in spring 2007 a commodities crisis; in mid 2008 the credit and banking crisis. The banking crisis had a huge effect on equity market volatility, with the VFTSE 30 reaching unprecedented levels of over 80%. Similar time-evolutions of the VSR term structures based on the other two formulae are not shown but are available on request.

Figure 5: VFTSE_E (Exchange's Methodology)



Note: The VFTSE term structure for maturities from 30 to 270 days between 12 October 1992 and 17 March 2009.

From henceforth, to implement (3) we adopt the numerical methods used by exchanges when they employ its discretisation (10) for quoting volatility indices, i.e. Riemann sum integration and linear interpolation. To implement (5) and (8) we employ cubic spline integration and Hermite cubic polynomial interpolation, as these have less numerical error. The three swap rates are denoted VSR(3), VSR(5) and VSR(8), respectively. Note that the main difference between VSR(3) and VSR(5) is the different numerical methods employed, whilst the difference between VSR(5) and VSR(8) is mostly due to the different theoretical formulas, and different discrete-monitorization biases.

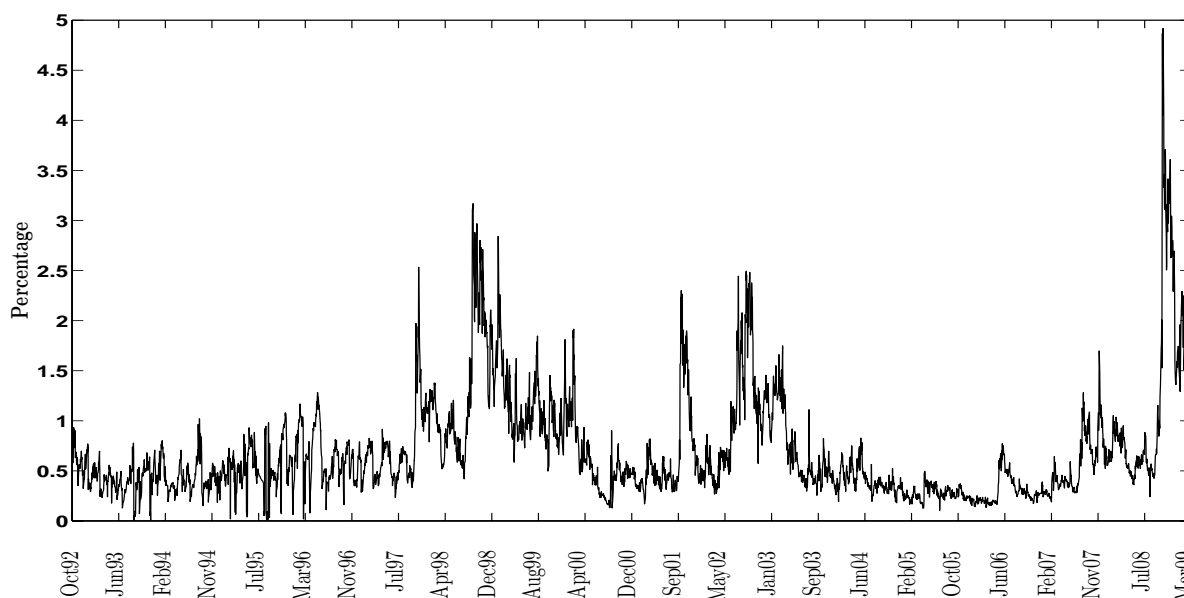
Table 2: VFTSE Descriptive Statistics (October 1992 to March 2009)

	30	60	90	120	150	180	210	240	270
VSR (3)									
Mean	20.409	20.326	20.288	20.450	20.645	20.594	20.464	20.331	20.197
StdDev	9.120	8.480	7.579	7.452	7.188	6.933	6.727	6.727	6.551
VSR (5)									
Mean	20.518	20.508	20.430	20.581	20.771	20.645	20.429	20.238	20.076
StdDev	9.455	8.842	8.184	7.963	7.887	7.651	7.413	7.197	6.970
VSR (8)									
Mean	19.694	19.479	19.293	19.263	19.276	19.118	18.912	18.716	18.531
StdDev	8.666	7.8612	7.203	6.890	6.728	6.477	6.242	6.052	5.893

Table 2 reports the mean and standard deviation of the daily VSRs of different maturities, computed as described above, over the period October 1992 to March 2009. Equation (5) yields the highest and most variable rates on average, although their overall characteristics are very close to those of the swap rates based on (3). As anticipated VSR(8) is less variable than the other two. It also tends to be lower on average; in fact, it is almost always lower than the other two rates. To illustrate this, consider Figure 6, which depicts the daily differences between the 30-day VSRs computed using the exchanges standard methodology and that based on (8) and implemented using the almost exact formula (12) for integration. The difference has been positive on every day since March 1996, and before that any negative difference was extremely small. Most days during tranquil periods, such as 2003-6, the difference is seldom more than 50 basis points. However, during the crises periods identified above the exchange's methodology

for quoting a 30-day VSR produced a result that was about 200 – 500 basis points greater than the VSR based on our implementation of (8).

Figure 6: VSR(3) – VSR(8)



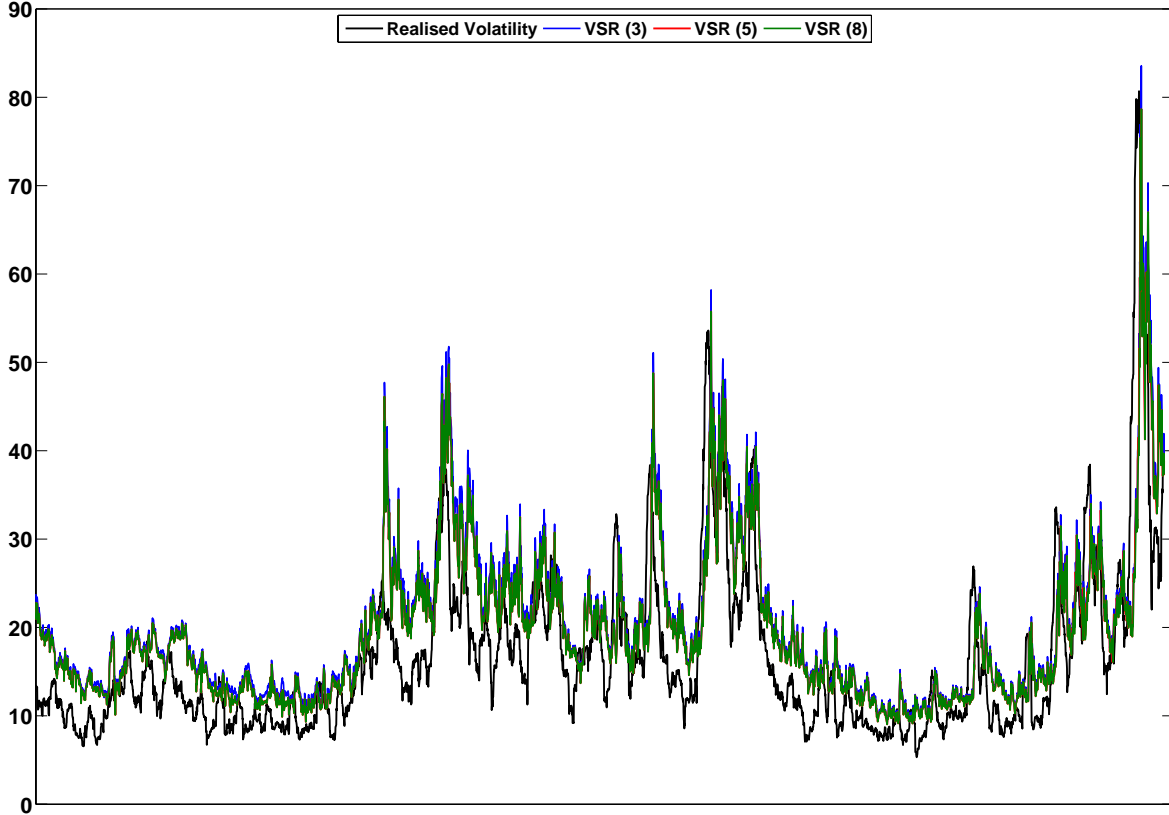
Note: Difference between the standard VFTSE (exchange’s methodology) and the VFTSE 30 based on formula (8) with cubic spline integration via (12)

Figure 7 shows the 30-day, discretely monitored realised volatility (in black) and the three 30-day VSRs, computed using (3), (5) and (8), the former using the exchange’s implementation and the two latter using Hermite spline variance term structure interpolation and cubic spline integration. The variance risk premium (the difference between the 30-day realised variance and the fair VSR) is negative during normal market circumstances, so banks should make profits from striking variance swaps at any of the VSRs. However, realised volatility jumps up before the crisis periods listed above, and the variance risk premium then became large and positive, indicating that substantial losses may have been incurred by variance swap dealers if they had not fully hedged their positions.

Which of the three swaps rates provides the most reliable prediction of the realised variance? In theory we should expect either (3) or (5) to be the most reliable predictors, because (8) relies on an approximation. But in practice, given the biases that we have discussed in this paper, the answer to this question is not clear.

For the three VSRs and for maturities 30, 60, . . . , 270 days, Table 3 reports descriptive statis-

Figure 7: Variance Swap Rates and Realised Volatility



Note: A comparison of the VSRs with the realised volatility between October 1992 and March 2009

tics of the variance risk premia, followed by the root mean square error (RMSE) of the VSR as a predictor of realised variance. Because (8) yields a lower VSR than the other two formulas it provides the best predictor of realised variance on average, for every maturity. The RMSEs for realised variance prediction based on VSR(8) are smaller than those based on (3) and the largest RMSEs occur when using the VSR (5). Since the numerical methods of cubic spline integration and Hermite spline interpolation were common to (5) and (8) it is the formula itself rather than its numerical implementation that drives the result.

6 Conclusions

This paper explores the following candidates for a fair-value of a variance swap on a semi-martingale process: (a) the formula used by exchanges, which is derived as the first moment of the quadratic variation of the log return but which is inaccurate in the presence of jumps, and which is implemented via linear interpolation over variance and Riemann sum integration, fol-

Table 3: Descriptive Statistics of Variance Risk Premium and RMSE Prediction Error

	30	60	90	120	150	180	210	240	270
VSR(3)									
Mean	-4.338	-4.010	-3.770	-3.625	-3.653	-3.576	-3.460	-3.371	-3.245
StDev	6.420	6.854	7.043	7.151	7.071	6.899	6.520	6.259	6.007
Min	-45.220	-44.605	-40.757	-34.247	-29.802	-26.285	-24.639	-22.219	-20.504
Max	29.013	25.377	22.728	21.567	22.506	22.018	20.531	19.570	18.560
RMSE	7.747	7.940	7.988	8.016	7.958	7.770	7.399	7.108	6.827
VSR(5)									
Mean	-4.444	-4.188	-3.913	-3.747	-3.759	-3.608	-3.407	-3.261	-3.113
StDev	6.534	6.994	7.175	7.301	7.213	7.047	6.683	6.426	6.152
Min	-30.095	-27.183	-23.434	-22.873	-23.794	-23.258	-21.661	-20.519	-19.267
Max	44.910	44.330	40.477	34.029	29.641	26.184	24.529	22.102	20.434
RMSE	7.901	8.151	8.172	8.205	8.133	7.916	7.519	7.205	6.894
VSR(8)									
Mean	-3.627	-3.173	-2.786	-2.469	-2.334	-2.150	-1.951	-1.795	-1.618
StDev	6.255	6.653	6.859	6.954	6.879	6.706	6.320	6.058	5.817
Min	-46.096	-45.396	-41.835	-35.513	-31.248	-27.837	-26.280	-23.947	-22.306
Max	27.409	21.346	19.899	18.357	19.161	18.552	16.736	15.781	15.395
RMSE	7.230	7.370	7.403	7.378	7.263	7.042	6.626	6.317	6.037

Note: This table presents some descriptive statistics and the root mean squared errors (RMSE) of the differences between the realised variance and the variance swap rate for all maturities up to 270 days between October 1992 and March 2009. All numbers are reported in annualized percentages

lowing the approach used by exchanges quoting volatility indices; (b) the formula first derived by Bakshi et al. [2003] and then generalized by Rompolis and Tzavalis [2009], based on the second moment of the log return, which we have implemented using Hermite spline interpolation over variance and numerical cubic spline integration; and (c) an approximate formula for the VSR derived from the risk-neutral price density and which we implemented using Hermite spline interpolation over variance and analytic cubic spline integration. Substantial empirical differences in these swap rates may be observed during turbulent market periods. For instance, during the banking crisis in the last quarter of 2008 some swap rates differed by 10 – 20 percentage points, depending on the theoretical formula and its numerical implementation.

Method (b) yields slightly larger swap rates than method (a), except during excessively volatile periods. At such times Riemann sum integration induces an upward bias on the result of method (a) which is only partially offset by the negative discrete-monitorization and jump biases inherent in the formula. By contrast, the discrete-monitorization bias in method (c) is smaller and could even be positive during very volatile periods. Yet the discrete-monitorization bias is relatively small compared with numerical implementation biases, and the swap rate (c)

is almost always lower than those based on (a) and (b).

The swap rates based on (c) are the most stable over time because both (a) and (b) require the weighting of option prices by the inverse of the square of their strike, making the resulting integration more sensitive to prices of low strike options. Often very low strike options are not actively traded so their close of day prices can be stale, and weighting them by the square of the strike enhances the bias due to option mispricing. Method (c) is also amenable to a fast and almost exact analytic integration over strike which has much smaller discretization bias than the standard Riemann sum approach. Empirically, although (c) is a formula for $\exp(\text{RV}) - 1$ rather than RV , the continuously-monitored realised variance over the life of the swap, the swap rates (c) have the lowest RMSE of prediction for the discretely-monitored realised variance based on over 16 years of daily vanilla option price data for the FTSE 100 index. Another advantage of the variance swap framework introduced in this paper is that it is derived from a risk-neutral price density rather than from moments of log returns. As such it is directly applicable to pricing interest rate variance swaps. Interest rates are already returns, and they can be zero or even negative, so the concept of a return on an interest rate is meaningless; realised variance for interest rate variance swaps should be defined on interest rate changes.

References

- G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16(1):101–143, 2003.
- D. Breeden and R. Litzenberger. State contingent prices implicit in option prices. *Journal of Business*, 51:621–651, 1978.
- P. Carr and R. Lee. Robust replication of volatility derivatives. *Unpublished paper: Courant Institute, NYU*, 2003.
- P. Carr and R. Lee. Volatility derivatives. *Annual Review of Financial Economics*, 1:1–21, 2009.
- P. Carr and D. Madan. Towards a theory of volatility trading. page 458. Cambridge Univ Pr, 2001.
- J. Clark. Return to variance? *Risk Magazine*, 23(2):online, 2010.
- C. De Boor. *A Practical Guide to Splines*. Springer Verlag, 2001.
- K. Demeterfi, E. Derman, M. Kamal, and J. Zou. A guide to volatility and variance swaps. *Journal of Derivatives*, 6(4):9–32, 1999.
- F. Fritsch and R. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 18:238–246, 1980.
- P. Friz and J. Gatheral. Pricing volatility derivatives as inverse problem. *Quantitative Finance*, 5(6):531–542, 2005.
- J. Gatheral. *The Volatility Surface: A Practitioner’s Guide*. Wiley, 2006.
- G.J. Jiang and Y.S. Tian. The model-free implied volatility and its information content. *Review of Financial Studies*, 18(4):1305, 2005. ISSN 0893-9454.
- G.J. Jiang and Y.S. Tian. Extracting model-free volatility from option prices: An examination of the VIX index. *Journal of Derivatives*, 14(3):35, 2007. ISSN 1074-1240.
- MorganMarkets. Variance Swaps. *European Equity Derivatives Research*, 2006.
- L. Rompolis and E. Tzavalis. Retrieving risk-neutral moments from option prices. *SSRN Working Paper*, (1), 2009.