Realized Volatility and Variance: Options via Swaps

Peter Carr* and Roger Lee[†]

This version: October 26, 2007

In this paper we develop strategies for pricing and hedging options on realized variance and volatility. Our strategies have the following features.

- Readily available inputs: We can use vanilla options as pricing benchmarks and as hedging instruments. If variance or volatility swaps are available, then we use them as well. We do not need other inputs (such as parameters of the instantaneous volatility dynamics).
- Comprehensive and readily computable outputs: We derive explicit and readily computable formulas for prices and hedge ratios for variance and volatility options, applicable at all times in the term of the option (not just inception).
- Accuracy and robustness: We test our pricing and hedging strategies under skew-generating volatility dynamics. Our discrete hedging simulations at a one-year horizon show mean absolute hedging errors under 10%, and in some cases under 5%.
- Easy modification to price and hedge options on *implied* volatility (VIX).

Specifically, we price and hedge realized variance and volatility options using variance and volatility swaps. When necessary, we in turn synthesize volatility swaps from vanilla options by the Carr-Lee [4] methodology; and variance swaps from vanilla options by the standard log-contract methodology.

1 Introduction

Let S_t denote the value of a stock or stock index at time t.

Given a variance/volatility option to be priced and hedged, let us designate as time 0 the start of its averaging period, and time T the end. For $t \in [0,T]$, and $\tau \leq t$, let $R^2_{\tau,t}$ denote the realized variance of returns over the time interval $[\tau,t]$.

The mathematical results about the synthesis of volatility and variance swaps will hold exactly if \mathbb{R}^2 refers to the continuously-monitored variance. This means the quadratic variation of $\log S$,

^{*}Bloomberg L.P. and NYU Courant Institute pcarr@nyc.rr.com.

[†]University of Chicago. RL@math.uchicago.edu. We thank the International Securities Exchange, Kris Monaco, and two anonymous referees for their very helpful advice and support.

times a constant conversion factor u^2 that includes any desired annualization and/or rescaling:

$$u^{2} \lim \sum_{\tau < t_{n} < t} \left(\log \frac{S_{t_{n}}}{S_{t_{n-1}}} \right)^{2} \tag{1.1}$$

where lim denotes limit in probability as the mesh of the partition $\{t_1 < t_2 < \cdots\}$ tends to zero. For example, choosing $u = 100 \times \sqrt{1/T}$ expresses $R_{0,T}^2$ in bp per unit time.

For practical implementation with daily monitoring, however, let N be the number of days in the period [0, T] between the daily closing times 0 and T, and let

$$R_{\tau,t}^2 := u^2 \sum_{\tau < t_n \le t} \left(\log \frac{S_{t_n}}{S_{t_{n-1}}} \right)^2 \tag{1.2}$$

where $t_0 := \tau$ and the $t_1 < t_2 < \dots$ are the successive daily closing times in $(\tau, t]$, together with t itself; and u is a constant annualization/rescaling factor. For example, choosing

$$u := 100 \times \sqrt{252/N} \tag{1.3}$$

expresses $R_{0,T}^2$ in units of annual bp.

The t will denote the valuation date; we allow arbitrary $t \in [0, T]$, because we will solve for prices and hedges not just at inception, but throughout the term of the option. The τ will denote the start date of the variance and volatility swaps that will serve as pricing benchmarks and hedging instruments; by not constraining $\tau = 0$, we have freedom to use swaps whose averaging periods $[\tau, T]$ need not coincide with the option's period [0, T].

For a given variance or volatility option, the u is fixed. It may depend on T, but T is fixed. It does not depend on t. Thus the $R_{\tau,t}^2$ is a scale factor times a running "cumulative" variance – not a running "average" variance, because the scale factor is designed to give a proper average only for the full interval [0,T].

Now consider swaps and options written on R^2 and $R := \sqrt{R^2}$.

A genuine $[\tau, T]$ variance swap with fixed leg f^2 pays the holder some notional amount times

$$R_{\tau,T}^2 - f^2. (1.4)$$

A genuine $[\tau, T]$ volatility swap with fixed leg f pays the holder some notional amount times

$$R_{\tau,T} - f. \tag{1.5}$$

In this paper, all swaps have notional 1 and fixed leg 0, unless otherwise stated.

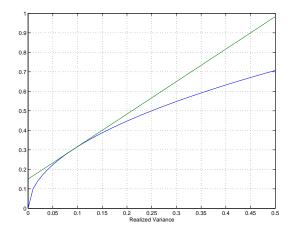
Our options treatment focuses on calls; results for puts can be obtained from parity relations. A [0,T] variance call with strike K_{vol}^2 pays the holder at time T some notional amount times

$$(R_{0,T}^2 - K_{\text{vol}}^2)^+. (1.6)$$

A [0,T] volatility call with strike K_{vol} pays the holder at time T some notional amount times

$$(R_{0,T} - K_{\text{vol}})^+$$
. (1.7)

Figure 2.1: Payoff of a volatility swap, compared to cash plus variance swaps



In this paper all options have notional 1, unless otherwise stated. The strike K_{vol} may be any nonnegative number. We say that a [0,T] variance option struck at K_{vol}^2 is at-the-money at time t if a [0,T] variance swap (on the same underlier) with fixed leg K_{vol}^2 has time-t value zero.

This paper proposes a methodology to price and hedge options on realized variance and volatility. Specifically, we price variance options using an explicit formula that takes as inputs the prices of variance and volatility swaps. We hedge variance options by trading variance and volatility swaps. We do likewise for volatility options.

If variance and volatility swaps are unavailable to trade, then we propose to *synthesize them* using vanilla options. So we begin with swaps, and build toward options.

2 Variance and volatility swaps

Variance option prices depend on the expectation and volatility of variance. The expectation is revealed by variance swap prices, and the volatility can be inferred from variance and volatility swap prices together. Specifically:

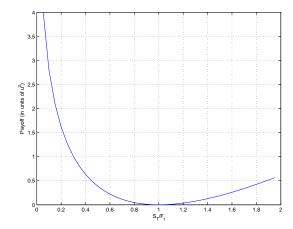
Let A_t be the time-t value of the variance swap which pays $R_{0,T}^2$.

Let B_t be the time-t value of the volatility swap which pays $R_{0,T}$.

Let r be the assumed constant interest rate, and let $A_t^* := A_t e^{r(T-t)}$ and $B_t^* := B_t e^{r(T-t)}$ be the time-t variance swap rate and volatility swap rate respectively; by definition this is the "fair" fixed leg, such that the variance swap (respectively volatility swap) has time-t value zero.

As shown in Figure 2.1, the volatility swap's concave square-root payoff is dominated by the linear payoff consisting of $\sqrt{A_t^*}$ in cash, plus $1/(2\sqrt{A_t^*})$ variance swaps with fixed leg A_t^* . The dominating payoff has forward value $\sqrt{A_t^*}$, because the variance swaps have value zero. Thus we have enforced Jensen's inequality $\sqrt{A_t^*} \geq B_t^*$ by superreplication.

Figure 2.2: The initial payoff profile of a synthetic variance swap



This concavity's price impact – as measured by how much $\sqrt{A_t^*}$ exceeds B_t^* – depends on the volatility of volatility. More precisely, letting \mathbb{E} denote risk-neutral expectation,

$$A_t^* - (B_t^*)^2 = \mathbb{E}_t R_{0,T}^2 - (\mathbb{E}_t R_{0,T})^2 = \text{Var}_t R_{0,T}.$$
(2.1)

So if we can obtain the swap values A and B, then we can back out the volatility-of-volatility, and use it to price options on $R_{0,T}^2$ and $R_{0,T}$. If we can moreover trade the variance and volatility swaps, then we can hedge the volatility-of-volatility risk. Similar reasoning holds for volatility options.

In the absence of genuine variance or volatility swaps, we obtain A and B from synthetic swaps.

2.1 Synthetic variance swap

By the theory developed in Neuberger [12], Dupire [8], Carr-Madan [5], and Derman et al [7], who assume essentially only the positivity and continuity of price paths, the following self-financing trading strategy replicates the continuously-monitored $R_{\tau,t}^2$ for a non-dividend-paying asset.

Write $F_t := S_t e^{r(T-t)}$ for the forward price, and choose an arbitrary put-call separator $\kappa > 0$. At each time $t \ge \tau$, hold the following static position in options, and dynamic position in shares:

$$u^{2} \frac{2}{K^{2}} dK \quad \text{calls at strikes } K > \kappa, \text{ puts at strikes } K < \kappa$$

$$u^{2} \left(\frac{2}{F_{t}} - \frac{2}{\kappa}\right) \quad \text{shares}$$

$$e^{-r(T-t)} \left[R_{\tau,t}^{2} + 2u^{2} \log(F_{t}/\kappa)\right] \quad \text{cash}$$

$$(2.2)$$

where all options have expiry T. We call this portfolio a *synthetic* variance swap. Its initial payoff profile appears in Figure 2.2. With continuous trading and a continuum of strikes, the final portfolio value will match the continuously-monitored variance.

2.2 Volatility swap: valuation under independent volatility

If one desires only to know the $[\tau, T]$ volatility swap's initial $t = \tau$ value (not the full replicating strategy), and if one assumes that instantaneous volatility evolves independently of the risk that drives price moves, then the at-the-money-forward implied volatility $\sigma_{\text{imp}}(F_t)$ approximates the desired volatility swap rate B_t^* , as the following argument shows.

Brenner-Subrahmanyam [2] found, by a Taylor expansion of the normal CDF about 0, that the at-the-money-forward Black-Scholes formula with volatility parameter σ satisfies

$$C^{BS}(\sigma) \approx \frac{S_t \sigma \sqrt{T - t}}{\sqrt{2\pi}}.$$
 (2.3)

Applying this twice,

$$\frac{S_t \sigma_{\text{imp}}(F_t) \sqrt{T - t}}{\sqrt{2\pi}} \approx C^{BS}(\sigma_{\text{imp}}(F_t)) = \mathbb{E}_t C^{BS}(R_{t,T}) \approx \mathbb{E}_t \frac{S_t R_{t,T} \sqrt{T - t}}{\sqrt{2\pi}} = \frac{S_t B_t^* \sqrt{T - t}}{\sqrt{2\pi}}$$
(2.4)

where the first equality holds due to the independence assumption. Therefore

$$B_t^* \approx \sigma_{\rm imp}(F_t) \tag{2.5}$$

as Feinstein [9] observed.

However, this estimate does not establish a replication strategy, does not apply at times $t > \tau$ after inception, and does not suggest how to handle the important case of correlated volatility. Our approach in section 2.3 addresses all of these issues.

2.3 Synthetic volatility swap: the Carr-Lee approach

The conventional wisdom holds that the pricing and hedging of a *volatility swap* is, unlike variance swaps, highly model-dependent.

The paper [4] of Carr-Lee challenges this notion. Without imposing a model on the dynamics of volatility, it shows how to replicate volatility swaps by trading vanilla options. It begins by making the independence assumption described in section 2.2 – but then it produces strategies robust to correlation. It also assumes frictionless trading in vanilla options, ignoring transactions costs – but these may be mitigated by the fact that only the *net* exposures in a portfolio need to be hedged via trades.

To specify the correlation-robust replication strategy, let I_{ν} denote the modified Bessel function of order ν – for which numerical implementations are readily available; for example, see Matlab's besseli or Mathematica's Besseli.

Figure 2.3: Initial payoff profile of a synthetic volatility swap: $R_{\tau,t}^2=0$

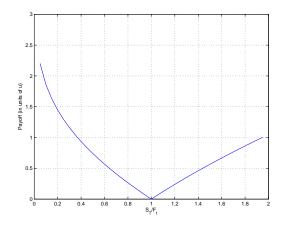


Figure 2.4: Evolution of the synthetic volatility swap: $R_{\tau,t}^2 = 0.25u^2$

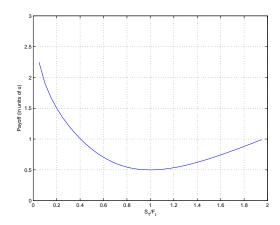
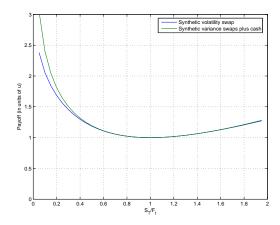


Figure 2.5: Evolution of the synthetic volatility swap: $R_{ au,t}^2=1.0u^2$



Initially, starting at time $t = \tau$, when $R_{\tau,t}^2 = 0$, the replicating portfolio holds

$$u \times \sqrt{\pi/2}/F_{\tau}$$
 straddles at strike $K = F_{\tau}$

$$u \times \sqrt{\frac{\pi}{8K^3F_{\tau}}} \left[I_1(\log \sqrt{K/F_{\tau}}) - I_0(\log \sqrt{K/F_{\tau}}) \right] dK \quad \text{calls at strikes } K > F_{\tau}$$

$$u \times \sqrt{\frac{\pi}{8K^3F_{\tau}}} \left[I_0(\log \sqrt{K/F_{\tau}}) - I_1(\log \sqrt{K/F_{\tau}}) \right] dK \quad \text{puts at strikes } K < F_{\tau}$$

$$(2.6)$$

together with a zero-cost delta-hedge. Afterwards, at times $t > \tau$ when $R_{\tau,t}^2 > 0$, it holds

$$u \times \frac{dK}{\sqrt{\pi}} \int_0^\infty \frac{e^{-zR_{\tau,t}^2/u^2}}{K^2 z^{1/2}} \left[\theta_+(K/F_t)^{p_+} + \theta_-(K/F_t)^{p_-} \right] dz \qquad \text{calls at strikes } K > F_t$$

$$puts at strikes K < F_t$$

$$e^{-r(T-t)} R_{\tau,t} \quad \text{cash}$$

$$(2.7)$$

together with a zero-cost delta-hedge, where

$$\theta_{\pm} := \frac{1}{2} \mp \frac{1}{2\sqrt{1-8z}} \qquad p_{\pm} := \frac{1}{2} \pm \sqrt{1/4 - 2z}.$$
 (2.8)

and all options have expiry T. The portfolio holdings at strike K depend only on the observables K, F, $R^2_{\tau,t}$, and u.

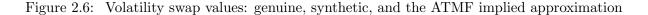
Under the independence assumption, with continuous trading in a continuum of strikes, we show in [4] that the portfolio self-finances and has time-T value equal to the continuously-monitored volatility swap payoff $R_{\tau,T}$. We call this portfolio a *synthetic* volatility swap.

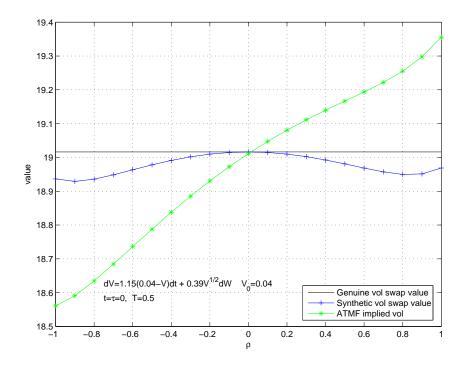
Figure 2.3 plots the initial payoff profile of the synthetic volatility swap, and the Appendix [3] implements this strategy using a discrete set of strikes, and gives a numerical example.

Most of the synthetic volatility swap's value resides in the ATMF straddles in (2.6). The short out-of-the-money calls and long out-of-the-money puts in (2.6) are precisely chosen to gain robustness to violations of the independence condition, by neutralizing the first-order effect of price/volatility correlation on the synthetic swap's value. This robustness to correlation is important in typical equity markets, where downward-sloping volatility skews indicate the presence of negative correlation.

Let us test how accurately the synthetic volatility swap value matches the genuine volatility swap value, under Heston [11] dynamics, with various correlation parameters $\rho \in [-1, 1]$. For comparison, we also include the at-the-money Black-Scholes implied volatility, as motivated by (2.5). Figure 2.6 shows the results for $t = \tau = 0$ and T = 0.5 and r = 0, using parameters estimated by Bakshi-Cao-Chen (BCC) [1], and $u = 100\sqrt{1/T}$. We compute the genuine volatility swap value using the identity $2\sqrt{\pi}\mathbb{E}R_{0,T} = \int_0^\infty (1 - \mathbb{E}e^{-zR_{0,T}^2})z^{-3/2}dz$ and the known Laplace transform of $R_{0,T}^2$.

Our synthetic volatility swap clearly outperforms implied volatility as an approximation of the genuine volatility swap value, across essentially all correlation assumptions: in the case $\rho = 0$, our method is (as promised) exact and the implied volatility approximation is nearly exact; but more importantly, in the empirically relevant case of $\rho \neq 0$, our synthetic volatility swap's relative "flatness" with respect to correlation results in its greater accuracy.





Taking $r \neq 0$ would change nothing, except to scale the genuine and synthetic volatility swap values by e^{-rT} . Taking T smaller (larger) tends to improve (worsen) the accuracy of both approximations: our synthetic volatility swap / the naive ATM-implied rule. For example, with $\rho = -0.64$, the respective -6/-30 bp discrepancies at T = 0.5 shown in Figure 2.6 would become -1/-13 bp at T = 0.25, and -18/-57 bp at T = 1.0. At each T, our synthetic volatility swap still has clearly greater accuracy than the naive ATM implied volatility rule.

Remark 2.1. The synthetic volatility swap evolves dynamically. Expressions (2.6) and (2.7) make this precise, but here we give some intuition. Initially it resembles an at-the-money straddle. Its dynamics depend on two factors. First, as the spot moves, the "strike" of the "straddle" floats to stay at-the-money. Second, as the running variance $R_{\tau,t}^2$ accumulates, the "straddle" smooths out; indeed, only when $R_{\tau,t}^2 = 0$ does the straddle's kink literally exist.

Eventually the synthetic volatility swap resembles a position in cash plus synthetic variance swaps, as shown in Figures 2.4 and 2.5. Intuitively, as variance accumulates, we progress rightward in Figure 2.1. In that direction, the square-root function loses convexity, and becomes more linear. Thus the cash-plus-variance-swaps portfolio becomes not merely an upper bound, but indeed improves as an approximation to the volatility swap.

3 Variance and volatility options

3.1 Pricing of variance and volatility options

We have the following pricing problem.

At time t, where $0 \le t < T$, we observe the following:

- $R_{0,t}^2$, the running variance from time 0 to t, where $0 \le t < T$
- $R_{\tau,t}^2$, the running variance from some time τ to t, where $\tau \leq t$
- A_t , the time-t price of a (genuine or synthetic) variance swap, which pays $R_{\tau,T}^2$
- B_t , the time-t price of a (genuine or synthetic) volatility swap, which pays $R_{\tau,T}$
- K_{vol} , the strike (quoted as a volatility)
- r, the interest rate. Let $G_t := e^{r(T-t)}$ denote the associated discount factor's reciprocal.

We have allowed the reference swaps' start date τ , the option's start date 0, and the valuation date t, to be distinct or identical. The condition $\tau \leq t$ allows either spot-starting or seasoned volatility swaps.

We intend to solve for

- \bullet Variance call: The time-t value of a claim on $(R_{0,T}^2-K_{\mathrm{vol}}^2)^+$
- Volatility call: The time-t value of a claim on $(R_{0,T} K_{\text{vol}})^+$

Our solution will approximate the time-t conditional distribution of $R_{\tau,T}$ as a displaced lognormal. This extends the use of a lognormal distribution by Friz-Gatheral [10], who assume that the valuation date, swap start date, and option start date all coincide; moreover, they do not address hedging; and for estimation of the input B, they do not use correlation-robust volatility swaps.

Specifically, let us approximate as lognormal the time-t conditional distribution of "remaining volatility" $R_{\tau,T} - R_{\tau,t}$. The lognormal distribution has two parameters (mean and variance), which we calibrate to the given variance swap and volatility swap prices. Using the calibrated lognormal distribution, we solve for prices of volatility and variance options.

Thus we obtain the following explicit pricing formulas which, like the Black-Scholes formula, involve N, the normal CDF.

The displaced lognormal price of a K_{vol}^2 -strike variance call is

$$C_{\text{LN}}^{\text{var}} = G_t^{-1} \times \begin{cases} A_t G_t + R_{0,t}^2 - R_{\tau,t}^2 - K_{\text{vol}}^2 & \text{if } K_{\text{vol}}^2 \le R_{0,t}^2 \\ \mu_2 N(d_0) + 2R_{\tau,t} \mu_1 N(d_1) - (K_{\text{vol}}^2 - R_{0,t}^2) N(d_2) & \text{if } K_{\text{vol}}^2 > R_{0,t}^2 \end{cases}$$
(3.1)

where

$$\mu_1 := B_t G_t - R_{\tau,t} \tag{3.2}$$

$$\mu_2 := A_t G_t + R_{\tau,t}^2 - 2B_t G_t R_{\tau,t} \tag{3.3}$$

and

$$d_j := \frac{m_t - \log((K_{\text{vol}}^2 + R_{\tau,t}^2 - R_{0,t}^2)^{1/2} - R_{\tau,t})}{s_t} + (2 - j)s_t, \qquad j = 0, 1, 2, \tag{3.4}$$

and where the log remaining volatility $\log(R_{\tau,T} - R_{\tau,t})$ has time-t conditional mean m_t computable via

$$m_t = 2\log\mu_1 - \frac{1}{2}\log\mu_2 \tag{3.5}$$

and has time-t conditional variance s_t^2 (squared "vol of vol") computable, from the relative sizes of the variance swap rate and volatility swap rate, via

$$s_t^2 = \log \mu_2 - 2\log \mu_1. \tag{3.6}$$

The displaced lognormal price of a K_{vol} -strike volatility call is

$$C_{\rm LN}^{\rm vol} = G_t^{-1} \int_{-\infty}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left((R_{0,t}^2 + 2R_{\tau,t}e^{m_t + s_t z} + e^{2m_t + 2s_t z})^{1/2} - K_{\rm vol} \right)^+ dz. \tag{3.7}$$

In the case that $\tau = 0$, this simplifies to

$$C_{\text{LN}}^{\text{vol}} = G_t^{-1} \times \begin{cases} B_t G_t - K_{\text{vol}} & \text{if } K_{\text{vol}} \le R_{0,t} \\ \mu_1 N(d_1) - (K_{\text{vol}} - R_{0,t}) N(d_2) & \text{if } K_{\text{vol}} > R_{0,t} \end{cases}$$
(3.8)

where μ_i and d_i are obtained by taking $\tau = 0$ in (3.2)-(3.4).

The pricing formulas are easily computable functions of observable quantities. Moreover, the pricing distribution prices exactly correctly any variance/volatility option known to finish in-the-money, including the benchmark swaps (which are zero-strike calls), as well as any option whose strike is smaller than the running variance/volatility; using an undisplaced lognormal distribution would violate this latter property.

Remark 3.1. Consider a volatility option in the case $t = \tau = 0$ and $\mu_1 = K_{\text{vol}}$. Thus the valuation date is 0, the pricing benchmarks are [0, T] swaps, and the volatility option is ATM. Then by (3.8),

$$C_0^* := e^{rT} C_{\text{LN}}^{\text{vol}} = \mu_1 [N(s_0/2) - N(-s_0/2)] \approx \frac{B_0^* s_0}{\sqrt{2\pi}} = B_0^* \left[\frac{1}{\pi} \log(\sqrt{A_0^*}/B_0^*) \right]^{1/2}$$
(3.9)

so the forward value of the ATM volatility option has a particularly simple expression in terms of the volatility swap rate B^* and the variance swap rate A^* . Moreover, rearranging (3.9) gives

$$\sqrt{A_0^*}/B_0^* = e^{\pi(C_0^*/B_0^*)^2},\tag{3.10}$$

so the "convexity correction" $\sqrt{A_0^*}/B_0^*$ has a simple explicit monotonic relationship with the ATM volatility option value C_0^*/B_0^* (expressed relative to the volatility swap rate).

3.1.1 Tests

The lognormal distribution is empirically a plausible approximation, but to test robustness, suppose that instantaneous volatility follows Heston dynamics (which do not generate lognormal distributions for realized variance), with the same parameters as in Figure 2.6. Hence the genuine and synthetic volatility swap values are as plotted in that figure, and the genuine and synthetic variance swap values are both 20^2 .

We plot in Figures 3.1 and 3.2 the lognormal approximations (using genuine and using synthetic volatility swap prices) to a variance call price, in comparison to the true Heston variance call price. Figure 3.1 varies the strike K_{vol} , whereas Figure 3.2 varies the price-volatility correlation ρ . In each figure the conversion factor is $u = 100\sqrt{1/T}$.

Even with non-lognormal Heston dynamics, and even when using synthetic volatility swaps, the Figures reveal errors of less than 4 bps (relative to a true value of about 100.4 bps), across all correlation assumptions, in the important ATM 20 strike.

At OTM strikes such as 30 and 35, our approximation has significant error (if the true dynamics are Heston), but the error's sign is consistent with the relative thinness of the Heston variance distribution's tail, compared to the lognormal tail; indeed if the true distribution of R has right tail equal or thinner than lognormal, then the displaced lognormal approximation can be seen as a conservative estimate, prudent from the standpoint of a dealer selling an OTM variance call.

3.2 Hedging of variance and volatility options

We propose to hedge variance and volatility options by $trading\ variance\ swaps\ and\ volatility\ swaps$ – either synthetic or genuine. Specifically, in order to hedge against a short position (or replicate a long position) in one variance call, hold at each time t

$$\frac{\partial C_{\text{LN}}^{\text{var}}}{\partial A}$$
 variance swaps, and $\frac{\partial C_{\text{LN}}^{\text{var}}}{\partial B}$ volatility swaps.

and in order to hedge against a short position (or replicate a long position) in one volatility call, hold at each time t

$$\frac{\partial C_{\text{LN}}^{\text{vol}}}{\partial A}$$
 variance swaps, and $\frac{\partial C_{\text{LN}}^{\text{vol}}}{\partial B}$ volatility swaps,

where all partial derivatives are evaluated at $(A_t, B_t, R_{0,t}^2, R_{\tau,t}^2, K_{\text{vol}})$.

The hedge ratios have the following explicit formulas. In the case of variance calls,

$$\frac{\partial C_{\text{LN}}^{\text{var}}}{\partial A} = \begin{cases}
1 & \text{if } K_{\text{vol}}^2 \leq R_{0,t}^2 \\
N(d_0) + \mu_2 N'(d_0) \frac{\partial D_0}{\partial A} + 2R_{\tau,t} \mu_1 N'(d_1) \frac{\partial D_1}{\partial A} - \chi N'(d_2) \frac{\partial D_2}{\partial A} & \text{if } K_{\text{vol}}^2 > R_{0,t}^2
\end{cases}$$

$$\frac{\partial C_{\text{LN}}^{\text{var}}}{\partial B} = \begin{cases}
0 & \text{if } K_{\text{vol}}^2 \leq R_{0,t}^2 \\
2R_{\tau,t}[N(d_1) - N(d_0)] + \mu_2 N'(d_0) \frac{\partial D_0}{\partial B} + 2R_{\tau,t} \mu_1 N'(d_1) \frac{\partial D_1}{\partial B} - \chi N'(d_2) \frac{\partial D_2}{\partial B} & \text{if } K_{\text{vol}}^2 > R_{0,t}^2
\end{cases}$$
(3.11)

Figure 3.1: Call price approximations, against strike. Correlation ρ is fixed at -0.64. The ATM strike (the variance swap rate, quoted as a volatility) is 20, and expiry is 0.5 years. In the left-hand plot, the vertical axis is the call's dollar value; in the right-hand plot, the vertical axis is the call's Black implied volatility. Heston dynamics are as in Figure 2.6.

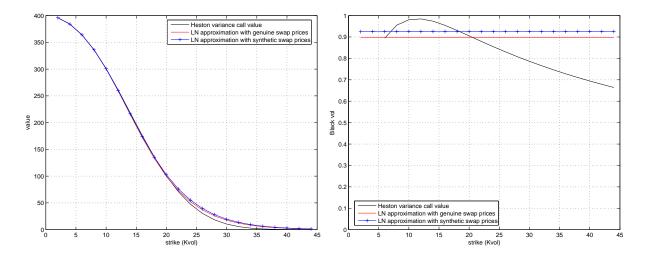
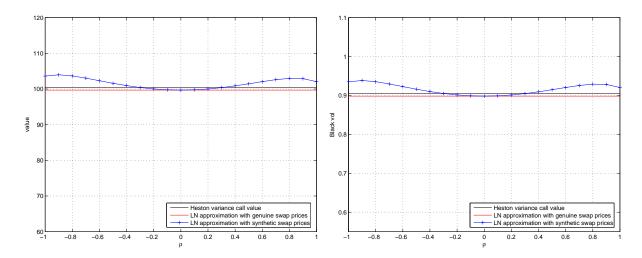


Figure 3.2: Call price approximations, against correlation. The strike is fixed at-the-money, and expiry is 0.5 years. In the left-hand plot, the vertical axis is the call's dollar value; in the right-hand plot, the vertical axis is the call's Black implied volatility. Heston dynamics are as in Figure 2.6.



where $\chi := K_{\text{vol}}^2 - R_{0,t}^2$ and

$$\frac{\partial D_j}{\partial B} := G_t^{-1} \frac{\partial d_j}{\partial B} = \frac{1}{s_t \mu_1} - \left(1 - j - \frac{m_t - \log((K_{\text{vol}}^2 + R_{\tau,t}^2 - R_{0,t}^2)^{1/2} - R_{\tau,t})}{s_t^2}\right) \left(\frac{1}{s_t \mu_1} + \frac{R_{\tau,t}}{s_t \mu_2}\right)$$
(3.13)

$$\frac{\partial D_j}{\partial A} := G_t^{-1} \frac{\partial d_j}{\partial A} = \frac{1}{2s_t \mu_2} \left(1 - j - \frac{m_t - \log((K_{\text{vol}}^2 + R_{\tau,t}^2 - R_{0,t}^2)^{1/2} - R_{\tau,t})}{s_t^2} \right)$$
(3.14)

for j = 0, 1, 2.

In the case of volatility calls with $\tau = 0$,

$$\frac{\partial C_{\text{LN}}^{\text{vol}}}{\partial A} = \begin{cases}
0 & \text{if } K_{\text{vol}} \leq R_{0,t} \\
\mu_1 N'(d_1) \frac{\partial D_1}{\partial A} - (K_{\text{vol}} - R_{0,t}) N'(d_2) \frac{\partial D_2}{\partial A} & \text{if } K_{\text{vol}} > R_{0,t}
\end{cases}$$
(3.15)

$$\frac{\partial C_{\text{LN}}^{\text{vol}}}{\partial B} = \begin{cases}
1 & \text{if } K_{\text{vol}} \leq R_{0,t} \\
N(d_1) + \mu_1 N'(d_1) \frac{\partial D_1}{\partial B} - (K_{\text{vol}} - R_{0,t}) N'(d_2) \frac{\partial D_2}{\partial B} & \text{if } K_{\text{vol}} > R_{0,t}
\end{cases}$$
(3.16)

For general τ , see the Appendix [3] equations (C.1) and (C.2).

The formulas of sections 3.1 and 3.2 allow arbitrary u, but assume notional 1. To apply the formulas to general notionals, just ensure that A and B are swap prices $per\ unit$ notional; then multiply the call price formulas by the call notional, and multiply the hedging formulas by the ratio of call notional to swap notional.

3.3 Implementation choices

To implement this strategy, the hedger must choose what type of variance and volatility swaps (genuine, synthetic, or a hybrid), and what start dates (fixed or rolling) to use for the swaps. Specifically, consider the following types of variance/volatility swaps.

- "Genuine": Use *genuine* swap quotes to infer the option's initial price and compute hedge ratios. Trade *genuine* swaps to implement the hedge.
- "Synthetic": Use *synthetic* swap quotes to infer the option's initial price and compute hedge ratios. Trade *synthetic* swaps to implement the hedge.
- "Hybrid": Use *genuine* swap quotes to set the option's initial price and compute hedge ratios.

 Trade *synthetic* swaps to implement the hedge.

Genuine quotes and genuine swap trades are desirable, but may not always be liquidly available, hence the importance of the synthetic alternatives.

All variance/volatility swaps will have terminal date T, but we have a choice of start dates τ .

• "Fixed": Use swaps which start at a fixed time τ . A natural choice is $\tau = 0$, which coincides with the start date of the option to be hedged.

• "Rolling": At each discrete time t_k when the hedge is rebalanced, use fresh swaps. Thus $\tau = t_k$ for the swaps that are held from time t_k until the next rebalancing time t_{k+1} .

For variance swaps, the fixed/rolling distinction is immaterial, by the additivity of variance; the dynamics of a variance swap that starts today differ only by a constant from one that started yesterday. For volatility swaps, however, the distinction matters, due to the square root.

In trading rolling-start synthetic volatility swaps, the bulk of the constituent transactions occur mainly among the most liquid vanilla options – opening a position in ATMF straddles, and closing a position in nearly-ATMF straddles; see Figure 2.3 and Appendix [3] Table B.1(c).

3.4 Simulation tests

To test these hedging strategies, we run the following simulation. Assume Heston dynamics with r=0 and BCC's [1] estimated parameters $\rho=-0.64$ and

$$dV_t = 1.15(0.20^2 - V_t)dt + 0.39\sqrt{V_t}dW_t, \qquad V_0 = 0.20^2$$
(3.17)

under risk-neutral measure. Under physical measure, assume that S has drift coefficient 0.06 and V has Heston dynamics with the same parameters as (3.17), except a long-run mean 0.18². As the Heston model does not capture all of the observed features of equity markets, the future research agenda includes supplementing these simulations with more sophisticated tests.

Let T = 1 year, and consider a variance call struck at-the-money at $K_{\text{vol}}^2 = 0.20^2$. Although we defer to future research any theoretical analysis of the possibly material effect of discrete sampling on variance option valuations, we do use in our simulation analysis the discretely (daily) sampled returns typical in practice.

Suppose we sell the option at the initial price inferred from (genuine or synthetic) volatility swap quotes. Then we hedge by trading (genuine or synthetic) variance and volatility swaps once per day, allowing synthetic volatility swaps to use vanilla options of all strikes. Define the hedging error to be the initial sale proceeds, plus the hedging P&L, minus the contractual option payout, defined discretely using daily returns.

We simulate 400 paths, and report in Table 3.3 the mean absolute hedging error for the variance call, under the strategy variations listed in Section 3.3. The errors are expressed as a percentage of the true initial option price. Figure 3.4 plots the distribution of the simulated error in hedging the variance call using rolling-start synthetic swaps.

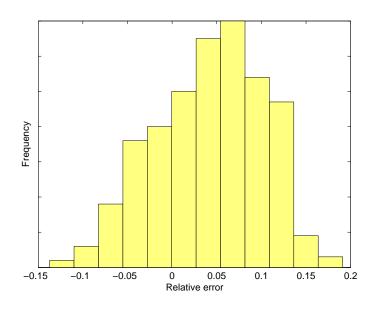
The rolling-start strategy outperforms the fixed-start strategy in these simulations. Intuitively, greater sensitivity to volatility-of-variance can be captured in a newly-issued volatility swap than in a seasoned volatility swap, making the former better suited to hedge volatility-of-variance.

The mean absolute hedging errors of under 10%, and under 5% for some strategies, indicate robustness of our methodology against unfavorable conditions, including the non-lognormal dynamics and nonzero price-correlation of the simulation's volatility process.

Table 3.3: Mean absolute errors of six variants of the hedging strategy for at-the-money 1-year variance calls, simulated under the Heston dynamics (3.17).

	Genuine swaps	Hybrid swaps	Synthetic swaps
Fixed-start swaps	4.7%	6.7%	8.1%
Rolling-start swaps	3.5%	5.5%	6.2%

Figure 3.4: Distribution of the hedging errors using rolling-start synthetic swaps in Table 3.3



4 Hedging VIX options

A special case of our setup, with modified inputs, produces prices and hedges of VIX options using variance swaps and VIX futures. Define $V_{0,T}^2 := \mathbb{E}_0 R_{0,T}^2$ to be the time-0 implied variance given by the time-0 variance swap rate. Define the [idealized] time-0 VIX to be $V_{0,T}$, where T = 1 month.

Consider a VIX call paying $(V_{0,T} - K_{\text{vol}})^+$ at time 0, which we will price and hedge at times $t \leq 0$. Redefine the discount factor $G_t := e^{r(0-t)}$, to reflect the pay date 0, not T. Redefine A_t as the time-t price of the payoff $V_{0,T}^2$, which admits replication by holding (when t < 0) variance swaps paying $R_{0,T}^2$ (and closing the position when t = 0); or alternatively replication via futures on $R_{0,T}^2$. Redefine B_t as the time-t price of the payoff $V_{0,T}$, which admits replication via VIX futures.

Define $R_{0,t} := R_{\tau,t} := 0$. Assuming the lognormality of $V_{0,T}$, the formula (3.8) prices the VIX option – in agreement with Carr-Wu [6] who, however, do not propose any hedge. Our formulas (3.15)-(3.16) propose how many units of A and B to hold, in a dynamic *hedge* of the VIX option.

More generally, assume the time-t conditional lognormality of $V_{0,T} - \beta$, for some nonegative displacement parameter $\beta < B_t$. Then, with $R_{0,t} := R_{\tau,t} := \beta$, our formulas (3.8, 3.15, 3.16) price and hedge VIX options. For $\beta > 0$, the resulting Black-implied volatility-of-VIX exhibits an upward skew ATM. We leave analysis of its accuracy to future research.

References

- [1] Gurdip Bakshi, Charles Cao, and Zhiwu Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, 52:2003–2049, 1997.
- [2] Menachem Brenner and Marti Subrahmanyam. A simple formula to compute the implied standard deviation. Financial Analysts Journal, 44(5):80–83, 1988.
- [3] Peter Carr and Roger Lee. Realized volatility and variance: Options via swaps. With Appendices. Download: http://math.uchicago.edu/~rl/OVSwithAppendices.pdf. Bloomberg LP and University of Chicago, 2007.
- [4] Peter Carr and Roger Lee. Robust replication of volatility derivatives. Download: http://math.uchicago.edu/~rl/rrvd.pdf. Bloomberg LP and University of Chicago, 2007.
- [5] Peter Carr and Dilip Madan. Towards a theory of volatility trading. In R. Jarrow, editor, *Volatility*, pages 417–427. Risk Publications, 1998.
- [6] Peter Carr and Liuren Wu. A tale of two indices. Journal of Derivatives, pages 13–29, 2006.
- [7] Emanuel Derman, Kresimir Demeterfi, Michael Kamal, and Joseph Zou. A guide to volatility and variance swaps. *Journal of Derivatives*, 6(4):9–32, 1999.
- [8] Bruno Dupire. Arbitrage pricing with stochastic volatility. Société Générale, 1992.
- [9] Steven P. Feinstein. The Black-Scholes formula is nearly linear in sigma for at-the-money options: Therefore implied volatilities from at-the-money options are virtually unbiased. Federal Reserve Bank of Atlanta, 1989.
- [10] Peter Friz and Jim Gatheral. Valuation of volatility derivatives as an inverse problem. *Quantitative Finance*, 5(6):531–542, 2005.
- [11] Steven L. Heston. A closed-form solution for options with stochastic volatility and applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.
- [12] Anthony Neuberger. Volatility trading. London Business School working paper, 1990.

A Appendix: Historical remarks

Variance swaps now trade actively over-the-counter. Like any swap, variance swaps are entered into at zero cost. Unlike most swaps, there is but a single exchange, which occurs at expiry. The buyer of a variance swap agrees to pay the difference between a standard ex-post calculation of realized variance and a fixed amount agreed upon at inception.

According to Mike Weber of Rabobank, the first variance swap appears to have been dealt in 1993 by him at the Union Bank of Switzerland (UBS). As the profile of the variance swap looked very much like that of an at-the-money-forward (ATMF) straddle, UBS initially valued the variance swap as such, less one vol point for safety. They later valued it using the method of Neuberger [12]. Mike Weber recalls that UBS bought one million pounds per vol point on the FTSE100 at 15 vol, with a cap at 21 (so UBS also dealt the first option on vol as well). The motivation for the trade was that UBS' book was short many millions of vega in the five-year time bucket and thus the trade represented a step in the right direction.

Volatility swaps also trade over-the-counter, but are not as liquid as variance swaps at present. Like a variance swap, a volatility swap is entered into at zero cost and involves a single payment at expiry. As the name suggests, the floating side of the payoff on a volatility swap is an expost measure of realized volatility, rather than variance. This volatility is obtained by taking the square root of the realized variance. For both variance and volatility swaps, the fixed payment is converted into a quote which is expressed in terms of annualized volatility.

Like many houses, UBS dealt a volatility swap before it did the variance swap in 1993. A casual historical survey suggests that most houses switched from initially dealing in volatility swaps to variance swaps, and some houses now deal in both. It is widely agreed that volatility swaps are harder to hedge in practice than variance swaps. This observation explains both the transition from vol swaps to variance swaps and the larger volume in the latter. Paralleling this transition, the definition of the Volatility Index (VIX) changed from an average of eight at-the-money implied volatilities to a weighted average of option prices in 2003. As this paper shows, this transition roughly corresponds to a change from a synthetic vol swap quote to a synthetic variance swap quote.

In the last few years, options on realized variance and volatility have also emerged on the scene. Like most over-the-counter options, the options are European-style and first appeared on stock indices. Besides the option embedded in the UBS 1993 variance swap, Mike Weber points out that during the late 1990s, several houses sold warrants with an embedded call on realized variance. However, with this kind of product, the option was typically struck far out-of-the-money (OTM). Similarly, the payoff to variance swaps on single names are always capped, thus embedding a short position in a deep-out-of-the-money call for the variance swap buyer. Nowadays, one can get quotes on at-the-money options on realized variance from several houses.

B Appendix: Implementation using discrete strikes

B.1 Synthetic volatility swap

First we¹ introduce some new notation to rewrite (2.6). Define $m := m(S, F) := \log(S/F)$ and

$$\psi(S,F) := u \times \sqrt{\frac{\pi}{2}} e^{m/2} |mI_0(m/2) - mI_1(m/2)|.$$
(B.1)

It can be shown that $\psi(S_T, F_\tau)$ is the payoff of the Carr-Lee synthetic $[\tau, T]$ volatility swap, and

$$\psi'(S, F) = u \times \frac{\text{sgn}(m)}{F_{\tau}} \sqrt{\frac{\pi}{2}} e^{-m/2} I_0(m/2), \qquad S \neq F,$$
 (B.2)

where the prime denotes partial derivative with respect to S. Moreover,

$$\psi''(S,F) = u \times \frac{\operatorname{sgn}(m)}{F_{\tau}^2} \sqrt{\frac{\pi}{8}} e^{-3m/2} (I_1(m/2) - I_0(m/2)), \qquad S \neq F.$$
 (B.3)

To construct a claim on the payoff $\psi(S_T, F_\tau)$ using a continuum of vanilla strikes, according to Carr-Madan [5], hold $\psi''(K, F_\tau)dK$ options at each strike K; in addition, because of the kink $\psi'(F_\tau \pm, F_\tau) = \pm u\sqrt{\pi/2}/F_\tau$, hold $u\sqrt{\pi/2}/F_\tau$ straddles at $K = F_\tau$. This agrees with (2.6).

However, when strikes are available only discretely, we replace dK by the strike spacing ΔK , and choose a highest put strike K_p and a lowest call strike K_c , where $K_p + \Delta K = K_c$. We recommend taking $K_p \leq F_{\tau} < K_c$, but our formulas will not assume this. Let K_* be the strike nearest F_{τ} (unless F_{τ} is equidistant between two strikes; then let $K_* := (K_p + K_c)/2$ to deactivate (B.7)).

The initial (time- τ) replicating portfolio (2.6) becomes

$$u\sqrt{\pi/2}/F_{\tau}$$
 straddles at strike $K = F_{\tau}$ (B.4)

$$\psi''(K, F_{\tau})\Delta K$$
 calls at strikes $K \geq K_c, K \neq K_*$ (B.5)

$$\psi''(K, F_{\tau})\Delta K$$
 puts at strikes $K \leq K_p, K \neq K_*$ (B.6)

$$\psi'(K_* + \Delta K/2, F_\tau) - \psi'(K_* - \Delta K/2, F_\tau) - u\sqrt{2\pi}/F_\tau \qquad \text{calls at strike } K_*, \text{ if } K_* \ge K_c \\ \text{puts at strike } K_*, \text{ if } K_* \le K_p \end{cases}$$
(B.7)

$$e^{-r(T-\tau)} \left[\frac{K_c - F_\tau}{K_c - K_p} \tilde{\psi}(K_p, F_\tau) + \frac{F_\tau - K_p}{K_c - K_p} \tilde{\psi}(K_c, F_\tau) \right] \quad \text{cash}$$
(B.8)

and a zero-cost delta hedge; see below. (B.9)

where $\tilde{\psi}(S,F) := \psi(S,F) - u\sqrt{\pi/2}|S/F - 1|$. We make the following line-by-line remarks.

The straddle at strike F_{τ} in (B.4) should be interpolated from the two available strikes K_0 and K_1 neighboring F_{τ} , where $K_0 \leq F_{\tau} < K_1$. For valuation purposes, each F_{τ} -straddle can be priced as $C^{BS}(\sigma_{\rm imp}(F_{\tau}))$, where C^{BS} is the Black-Scholes straddle (call plus put) formula, and $\sigma_{\rm imp}(F_{\tau})$ is the Black-Scholes implied volatility linearly interpolated between the observable implied volatilities at strikes K_0 and K_1 . For hedging purposes, the F_{τ} -strike straddle can be approximated

We thank Glen Luckjiff and Mark Shaps for correcting typos in a previous version of this Appendix.

as $(K_1 - F_\tau)/(K_1 - K_0)$ straddles struck at K_0 , plus $(F_\tau - K_0)/(K_1 - K_0)$ straddles struck at K_1 , plus the amount of cash needed to make the total value equal to $C^{BS}(\sigma_{imp}(F_\tau))$.

The synthetic volatility swap's remaining components (B.5) to (B.8) aim to match discretely the value of the payoff $\tilde{\psi}(S_T, F_\tau)$, which is ψ minus the straddle. The call and put quantities in (B.5) and (B.6) use the standard second derivative, but in (B.7) the K_* option quantity is specified as a finite difference, which more accurately deals with the ψ function's nonsmoothness at F_τ .

Whereas the options price the convexity of ψ , the cash position (B.8) prices the level and slope of ψ at the put-call-separating strike. To specify this cash position, we again use a finite-difference version of the standard formula, to deal with the nonsmoothness at F_{τ} . The cash position will typically be negligible if K_p and K_c are chosen close to F_{τ} .

The zero-cost delta-hedge in (B.9) does not affect valuation; but for hedging, it delta-neutralizes each option at each strike K using $-D_t(K)$ shares and $D_t(K)S_t$ in cash, where

$$D_{\tau}(K) := \text{Delta}^{BS}(\sigma_{\text{imp}}(K)) - \text{Vega}^{BS}(\sigma_{\text{imp}}(K)) \frac{K}{S_{\tau}} \frac{\partial \sigma_{\text{imp}}}{\partial K}$$
(B.10)

and Delta^{BS} and Vega^{BS} are the Black-Scholes delta and vega. Alternatively, $-D_{\tau}(K)e^{-r(T-\tau)}$ futures contracts (and no cash) may be used for each option at each strike K. Under the condition of price/volatility independence, the options position in (2.6) is already delta neutral, so the total share position automatically evaluates to zero, as shown in Carr-Lee [4]. Absent the independence condition, however, we have a (typically nonzero) total share position, which neutralizes the options' delta, robustly in that the only assumption is degree 1 homogeneity of option prices in spot and strike.

B.2 Numerical example: Volatility and variance swaps

At time t = 0, we construct synthetic [0, T] volatility and variance swaps on S with expiry T = 0.5. Suppose that $S_0 = 100$ and r = 0.04, and that T-expiry vanilla calls and puts on S are available at strike increments of $\Delta K = 5$.

Table B.1 performs the calculations, resulting in the synthetic volatility swap value $B_0 = 19.41$, and the synthetic variance swap value $A_0 = 395.09$, where the units of variance are annual basis points, as specified in (1.3), and the swaps have notional 1.

B.3 Numerical example: Volatility and variance options

In the setting of B.2, consider a [0,T] variance call and a [0,T] volatility call, each with strike $K_{\text{vol}} = 20$. Suppose volatility swaps and calls have notional 1, but the variance swaps and calls have notional $1/(2K_{\text{vol}}) = 1/40$ (which gives them the same "vega notional" as the volatility claims).

By (3.1) and (3.8), with the A and B inferred in section B.2, we find a variance call value of 1.34 and a volatility call value of 1.20. The variance call hedge consists of 2.93 variance swaps and -2.39 volatility swaps. The volatility call hedge consists of 2.35 variance swaps and -1.88 volatility swaps.

Table B.1: Synthetic volatility and variance swaps

Strike	Option type	Number of options		Premium per option	Delta per option
		Volatility swap	Variance swap		
(a)	(b)	(c)	(d)	(e)	(f)
60	put	0.109	55.56	0.02	-0.002
65	put	0.094	47.34	0.05	-0.005
70	put	0.083	40.82	0.11	-0.012
75	put	0.073	35.56	0.25	-0.026
80	put	0.065	31.25	0.51	-0.050
85	put	0.059	27.68	0.97	-0.091
90	put	0.053	24.69	1.74	-0.154
95	put	0.048	22.16	2.92	-0.244
100	put	0.036	20.00	4.65	-0.361
102.0	straddle	1.737	0	11.05	0.170
105	call	-0.040	18.14	4.10	0.500
110	call	-0.037	16.53	2.26	0.353
115	call	-0.035	15.12	1.07	0.216
120	call	-0.032	13.89	0.42	0.110
125	call	-0.030	12.80	0.12	0.044
130	call	-0.028	11.83	0.03	0.012

We have $F_0 = S_0 e^{rT} = 102.0$, so we take $K_p = 100$, $K_c = 105$.

Column (c) is computed by (B.4) through (B.7). Column (d) is computed by (2.2). Column (e) is observed from market prices of listed options; in this example, we suppose that the prices correspond to an implied volatility skew $\sigma_{\rm imp}(K) = 0.20 - 0.002(K - 100)$. Column (f) is by (B.10). Therefore we have

and the volatility swap's zero-cost delta hedge should have delta $-\sum[(c) \times (f)] = -0.204$, which may be implemented as -0.204 shares plus $0.204 \times S_0$ cash, or alternatively as $-0.204 \times e^{-rT}$ futures (and zero cash).

C Appendix: Hedge ratios

For a call on volatility, with general τ , we have the hedge ratios

$$\frac{\partial C_{\text{LN}}^{\text{vol}}}{\partial A} = \begin{cases}
0 & \text{if } K_{\text{vol}} \leq R_{0,t} \\
\int_{-d_2}^{\infty} \frac{R_{\tau,t} e^{m_t + s_t z} + e^{2m_t + 2s_t z}}{(R_{0,t}^2 + 2R_{\tau,t} e^{m_t + s_t z} + e^{2m_t + 2s_t z})^{1/2}} \left(\frac{z}{2s_t \mu_2} - \frac{1}{2\mu_2}\right) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz & \text{if } K_{\text{vol}} > R_{0,t} \end{cases}$$

$$\frac{\partial C_{\text{LN}}^{\text{vol}}}{\partial B} = \begin{cases}
1 & \text{if } K_{\text{vol}} \leq R_{0,t} \\
\int_{-d_2}^{\infty} \frac{R_{\tau,t} e^{m_t + s_t z} + e^{2m_t + 2s_t z}}{(R_{0,t}^2 + 2R_{\tau,t} e^{m_t + s_t z} + e^{2m_t + 2s_t z})^{1/2}} \left(\frac{2}{\mu_1} + \frac{R_{\tau,t}}{\mu_2} - \frac{z}{s_t \mu_1} - \frac{zR_{\tau,t}}{s_t \mu_2}\right) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz & \text{if } K_{\text{vol}} > R_{0,t} \end{cases}$$
(C.1)

by differentiating (3.7).