USC FBE FINANCE SEMINAR
presented by Ian Martin
FRIDAY, Sept. 16, 2011
10:30 am – 12:00 pm, **Room: JKP-202**

Simple Variance Swaps

Ian Martin*

September, 2011

Abstract

The events of 2008-9 disrupted volatility derivatives markets and caused the single-name variance swap market to dry up completely; it has never recovered. This paper introduces the *simple variance swap*, a relative of the variance swap with more desirable properties. Simple variance swaps are robust: they can be priced and hedged even if the underlying asset's price can jump. I construct SVIX, an index based on simple variance swaps that measures market volatility, and compare it to VIX. The SVIX series implies a lower bound on the forward-looking equity premium that peaked at 55% at the height of the credit crisis.

Keywords: variance swap, gamma swap, VIX, robust hedging, entropy, equity premium forecast, jumps, implied correlation, dispersion trade.

*Stanford GSB and NBER; ian.martin@gsb.stanford.edu; http://www.stanford.edu/~iwrm/. First draft: November 15, 2010. I thank Peter Carr, John Cochrane, Darrell Duffie, Stefan Hunt, Stefan Nagel, Myron Scholes, and participants at the SITE 2011 conference for their comments.

In recent years, a large market in volatility derivatives has developed. An emblem of this market, the VIX index, is often described in the financial press as "the fear index"; its construction is based on theoretical results on the pricing of variance swaps. These derivatives permit investors and dealers to hedge and to speculate in volatility itself. They also play an informational role by providing evidence about perceptions of future volatility.

The events of 2008 and 2009 severely disrupted these markets and revealed certain undesirable features of variance swaps. Carr and Lee (2009) write, "The cataclysm that hit almost all financial markets in 2008 had particularly pronounced effects on volatility derivatives .... Dealers learned the hard way that the standard theory for pricing and hedging variance swaps is not nearly as model-free as previously supposed .... In particular, sharp moves in the underlying highlighted exposures to cubed and higher-order daily returns. The inability to take positions in deep OTM options when hedging a variance swap later affected the efficacy of the hedging strategy. As the underlying index or stock moved away from its initial level, dealers found themselves exposed to much more vega than a complete hedging strategy would permit. This issue was particularly acute for single names, as the options are not as liquid and the most extreme moves are bigger. As a result, the market for single-name variance swaps has evaporated in 2009." Nor has it recovered subsequently. In response to this sensitivity to extreme events, market participants have imposed caps on variance swap payoffs. These caps limit the maximum possible payoff on a variance swap, at the cost of complicating the pricing and interpretation of the original contract.

In the first part of this paper, I define and analyze a financial contract that I call a *simple variance swap*. Although it has, arguably, a simpler definition than that of a standard variance swap,[1] I show that simple variance swaps are robust to the issues mentioned in the quotation above, and explain why. Simple variance swaps can be priced and hedged under far weaker assumptions than are required for pricing variance swaps (or the recently introduced gamma swaps). In particular, they can be hedged in the presence of jumps.

If, in the phrase of Summers (1985), this first part of the paper is "ketchup economics", then the second part adds some economic meat. The definition of the VIX index is based on the strike of a variance swap. I define the analogous index based on the strike of a simple variance swap, and call it SVIX. I prove various results about the properties of VIX and SVIX. In particular I show, under weak assumptions, that SVIX provides a lower bound

[1] I will refer to this standard product as a variance swap, and to the proposed alternative as a simple variance swap.

for the forward-looking equity premium. Using option price data from *OptionMetrics*, I construct the SVIX index and the implied equity premium lower bounds. The results are striking. Although the average lower bound in the data is on the order of 5%, there is extraordinary variation in the bound. The *lower bound* on the one-month equity premium peaked at 55.0% (in annualized terms) at the height of the 2008–9 crisis, and the lower bound on the one-year equity premium peaked at 21.5%.

Lastly, I consider the problem of computing market-implied correlations. In this respect, too, simple variance swaps are an improvement on variance swaps, both for observers interested in the market's view of implied correlation and for investors seeking to hedge or speculate in correlation.

Section 1 introduces the definition of a simple variance swap, and presents the main pricing and hedging results. Result 1 is a straightforward application of no-arbitrage logic. Results 2 and 3 derive an important simplification in the natural limiting case. Section 2 compares these results to those already known for standard variance swaps. Existing results are summarized in Result 4. These depend on the assumption that the underlying asset's price cannot jump. Next, I propose an index, SVIX, that is analogous to VIX but based on the strike on a simple variance swap rather than on a (standard) variance swap. Results 5, 6, and 7 show how VIX can be interpreted in the presence of jumps, and compare to the much simpler interpretation of the proposed SVIX index. Section 3 presents the lower bounds on the equity premium, and discusses the measurement of implied correlation. Section 4 concludes.

Related literature. The literature on variance swaps (Carr and Madan (1998), Demeterfi et al. (1999)) is based on papers by Breeden and Litzenberger (1978) and Neuberger (1990, 1994). Carr and Corso (2001) propose a contract related to the simple variance swap proposed below, but their pricing and hedging methodology is only valid if the underlying is a futures contract. Lee (2010a, 2010b) provides a brief summary of volatility derivative pricing in the absence of jumps. Finally, the paper of Carr and Lee (2009) quoted above provides an excellent survey of the state of the art in the area.

1 Pricing and hedging a simple variance swap

Consider an underlying asset whose price at time t is S_t. The interest rate is assumed constant, at r, throughout; this is the standard assumption in the related literature. The asset continuously pays dividends at rate δS_t per unit time. In subsequent sections, I will assume that the asset does not pay dividends, $\delta = 0$, since, again, this is the assumption generally made in the related literature; for now, we consider the general case because it comes at a low cost in terms of extra complexity but has obvious practical advantages. For example, if the underlying asset is a currency then the "dividend yield" δ corresponds to the foreign interest rate. Later in this section, and in the appendix, I extend to more general dividend payout policies, including the cases of completely unanticipated dividend payouts and of dividend payouts that are known at time 0.

We can now define a *simple variance swap* on this underlying asset. This is an agreement to exchange, at time T, some prearranged payment V (the "strike" of the simple variance swap) for the amount

$$\left(\frac{S_\Delta - S_0}{S_0}\right)^2 + \left(\frac{S_{2\Delta} - S_\Delta}{F_\Delta}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{T-\Delta}}\right)^2. \tag{1}$$

The quantities F_t that appear in the denominators are the forward prices to time t of the underlying asset; they are known at time 0. Given the above assumptions, $F_t = S_0 e^{(r-\delta)t}$. I call the contract a simple variance swap because (i) it is simple to hedge and price, even in the presence of jumps; and (ii) the strike V will turn out to measure the risk-neutral variance of the simple return of the underlying asset.

The pricing problem is to choose the strike V so that no money changes hands at initiation. The choice to put forward prices in the denominators is important: below we will see that this choice leads to a dramatic simplification of the formula for the strike V, and of the associated hedging strategy, in the limit as the period length Δ goes to zero. In an idealized frictionless market, this simplification of the hedging strategy would merely be a matter of analytical convenience; in practice, with trade costs, it acquires far more significance.

The following result shows how to price a simple variance swap (i.e. how to choose V so that no money need change hands initially) in terms of the prices of European call and put options on the underlying asset. I write $\mathsf{call}_t(K)$ for the time-0 price of a European call option on the underlying asset, expiring at time t, with strike K, and $\mathsf{put}_t(K)$ for the time-0 price of the corresponding put option.

Result 1 (Pricing)**.** *The strike on a simple variance swap, i.e. an agreement to exchange V for the amount in* (1) *at time T, is given by*

$$V(\Delta, T) = \sum_{i=1}^{T/\Delta} \frac{e^{ri\Delta}}{F_{(i-1)\Delta}^2} \left[\Pi(i\Delta) - (2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta}\Pi((i-1)\Delta)\right], \tag{2}$$

where $\Pi(t)$ *is the time-0 price of a claim to* S_t^2 *paid at time* t*, which is given by*

$$\Pi(t) = 2\int_0^{F_t} \mathsf{put}_t(K)\,dK + 2\int_{F_t}^{\infty} \mathsf{call}_t(K)\,dK + e^{-rt}F_t^2, \qquad \Pi(0) = S_0^2. \tag{3}$$

Proof. The absence of arbitrage implies that there exists a sequence of strictly positive stochastic discount factors $M_\Delta, M_{2\Delta}, \ldots$ such that a payoff $X_{j\Delta}$ at time $j\Delta$ has price

$$\mathbb{E}_{i\Delta}\left[M_{(i+1)\Delta}M_{(i+2)\Delta}\cdots M_{j\Delta}X_{j\Delta}\right]$$

at time $i\Delta$. The subscript on the expectation operator indicates that the expectation is conditional on time-$i\Delta$ information. I abbreviate $M_{(j\Delta)} \equiv M_\Delta M_{2\Delta}\cdots M_{j\Delta}$.

V is chosen so that the swap has zero initial value, i.e.,

$$\mathbb{E}\left[M_{(T)}\left\{\left(\frac{S_\Delta - S_0}{S_0}\right)^2 + \left(\frac{S_{2\Delta} - S_\Delta}{F_\Delta}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{T-\Delta}}\right)^2 - V\right\}\right] = 0. \tag{4}$$

We have

$$\begin{aligned}\mathbb{E}\left[M_{(T)}\left(S_{i\Delta} - S_{(i-1)\Delta}\right)^2\right] &= e^{-r(T-i\Delta)}\,\mathbb{E}\left[M_{(i\Delta)}\left(S_{i\Delta} - S_{(i-1)\Delta}\right)^2\right] \\ &= e^{-r(T-i\Delta)}\left\{\mathbb{E}\left[M_{(i\Delta)}S_{i\Delta}^2\right] - \left(2e^{-\delta\Delta} - e^{-r\Delta}\right)\mathbb{E}\left[M_{((i-1)\Delta)}S_{(i-1)\Delta}^2\right]\right\},\end{aligned}$$

using the law of iterated expectations; the fact that the interest rate r is constant, so that $\mathbb{E}_{(i-1)\Delta}\,M_{i\Delta} = e^{-r\Delta}$; and the fact that $\mathbb{E}_{(i-1)\Delta}\,M_{i\Delta}S_{i\Delta} = e^{-\delta\Delta}S_{(i-1)\Delta}$, which expresses the fact that if dividends are continuously reinvested, an investment of $e^{-\delta\Delta}S_{(i-1)\Delta}$ at time $(i-1)\Delta$ is worth $S_{i\Delta}$ at time $i\Delta$. If we define $\Pi(i)$ to be the time-0 price of a claim to S_i^2, paid at time i, then we have[2]

$$\mathbb{E}\left[M_{(T)}\left(S_{i\Delta} - S_{(i-1)\Delta}\right)^2\right] = e^{-r(T-i\Delta)}\left[\Pi(i\Delta) - \left(2 - e^{-(r-\delta)\Delta}\right)e^{-\delta\Delta}\Pi((i-1)\Delta)\right].$$

Equation (2) follows on substituting this into (4), so it only remains to confirm that $\Pi(t)$ is indeed given by (3). To see that it does, note that[3]

$$S_t^2 = 2\int_0^{\infty} \max\{0, S_t - K\}\,dK.$$

[2]Neuberger (1990), the working paper version of Neuberger (1994), mentions the relevance of this "Squared contract" in a similar context.

[3]Darrell Duffie suggested this approach. A previous draft derived (5) via Breeden-Litzenberger (1978) logic and integration by parts: $\Pi(t) = \int_0^\infty K^2\,\mathsf{call}_t''(K)\,dK = 2\int_0^\infty \mathsf{call}_t(K)\,dK$. The latter approach is less neat, but has the advantage of being mechanical: it does not rely on a "trick".

The right-hand side is the time-t payoff on a portfolio of European call options of all strikes, so the absence of arbitrage implies that

$$\Pi(t) = 2\int_0^\infty \mathsf{call}_t(K)\, dK. \tag{5}$$

Equation (3) follows by put-call parity, which is the relationship $\mathsf{call}_t(K) = \mathsf{put}_t(K) + e^{-rt}(F_t - K)$. Although (3) is less concise than (5), it has the appealing feature that it expresses $\Pi(t)$ in terms of out-of-the-money options only. $\square$

The most important aspect of Result 1 is that it does not require the price process of the underlying asset to be continuous. The strike on a simple variance swap is dictated by the prices of options across all strikes and the whole range of expiry times Δ, 2Δ, $\ldots$, T. But, correspondingly, the hedging portfolio requires holding portfolios of options of each of these maturities. Although this is not a serious issue if Δ is large relative to T, it raises the concern that hedging a simple variance swap may be extremely costly in practice if Δ is very small relative to T. Fortunately, this concern is misplaced: both the pricing formula (2) and the hedging portfolio simplify nicely in the limit as $\Delta \to 0$, holding T constant. I write

$$P(t) = 2\left\{\int_0^{F_t} \mathsf{put}_t(K)\, dK + \int_{F_t}^\infty \mathsf{call}_t(K)\, dK\right\}, \quad P(0) = 0.$$

Result 2 (Pricing, continued). *In the limit as $\Delta \to 0$, we have*

$$V(0,T) \equiv \lim_{\Delta\to 0} V(\Delta, T) = \frac{e^{rT}P(T)}{F_T^2}. \tag{6}$$

Proof. Observe that (2) can be rewritten

$$V(\Delta,T) = \sum_{i=1}^{T/\Delta}\left\{\frac{e^{ri\Delta}}{F_{(i-1)\Delta}^2}\left[P(i\Delta) - (2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta}P((i-1)\Delta)\right]\right\} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2. \tag{7}$$

For $0 < j < T/\Delta$, the coefficient on $P(j\Delta)$ in equation (7) is

$$\frac{e^{rj\Delta}}{F_{(j-1)\Delta}^2} - \frac{e^{r(j+1)\Delta}}{F_{j\Delta}^2}(2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta} = \frac{e^{rj\Delta}}{F_{j\Delta}^2}\left(e^{(r-\delta)\Delta} - 1\right)^2.$$

We can therefore rewrite (7) as

$$V(\Delta,T) = \frac{e^{rT}}{F_{T-\Delta}^2}P(T) + \underbrace{\sum_{j=1}^{T/\Delta - 1}\frac{2e^{rj\Delta}}{F_{j\Delta}^2}\left(e^{(r-\delta)\Delta} - 1\right)^2 P(j\Delta)}_{O(1/\Delta)\text{ terms of size }O(\Delta^2)} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2.$$

The second term on the right-hand side is a sum of $T/\Delta - 1$ terms, each of which has size on the order of Δ^2; all in all, the sum is $O(\Delta)$. The third term on the right-hand side is also $O(\Delta)$, so both tend to zero as $\Delta \to 0$. The first term tends to $e^{rT}P(T)/F_T^2$, as required. □

Result 2 is the motivation behind the choice of forward prices as the normalizing weights in the definition (1). In principle, we could have put any other constants known at time 0 in the denominators of the fractions in (1). Had we done so, we would have to face the unappealing prospect of a hedging portfolio requiring positions in options of all maturities between 0 and T. Using forward prices lets us sidestep this problem.[4]

The proof of Result 1 implicitly supplies the dynamic trading strategy that replicates the payoff on a simple variance swap. Tables 2 and 3 in the Appendix describe the strategy in detail. Each row of Table 2 indicates a sequence of dollar cashflows that is attainable by investing in the asset indicated in the leftmost column. Negative quantities indicated that money must be invested; positive quantities indicate cash inflows. Thus, for example, the first row indicates a time-0 investment of $\$e^{-rT}$ in the riskless bond maturing at time T, which generates a time-T payoff of \$1. The second and third rows indicate a short position in the underlying asset, held from 0 to Δ with continuous reinvestment of dividends, and subsequently rolled into a short bond position. The fourth row represents a position in a portfolio of call options of all strikes expiring at time Δ, as in equation (5); this portfolio has simple return $S_\Delta^2/\Pi(\Delta)$ from time 0 to time Δ. The fifth, sixth, and seventh rows indicate how the proceeds of this option portfolio are used after time Δ. Part of the proceeds are immediately invested in the bond until time T; another part is invested from Δ to 2Δ in the underlying asset, and subsequently from 2Δ to T in the bond. The replicating portfolio requires similar positions in options expiring at times $2\Delta, 3\Delta, \ldots, T - 2\Delta$. These are omitted from Table 2, but the general such position is indicated in Table 3, together with the subsequent investment in bonds and underlying that each position requires.

The self-financing nature of the replicating strategy is reflected in the fact that the total of each of the intermediate columns from time Δ to time $T - \Delta$ is zero. The last column of Table 2 adds up to the desired payoff,

$$\left(\frac{S_\Delta - S_0}{S_0}\right)^2 + \left(\frac{S_{2\Delta} - S_\Delta}{F_\Delta}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{T-\Delta}}\right)^2 - V.$$

[4] In the Appendix, I show that (6) also provides an excellent approximation to $V(\Delta, T)$ for $\Delta > 0$. Suppose that $r - \delta = 3\%$. If the intervals are daily, $\Delta = 1/252$, the error in the approximation is less than 0.01% of $V(\Delta, T)$; if the intervals are weekly, $\Delta = 1/52$, then it is less than 0.05%.

Therefore, the first column must add up to the cost of entering the simple variance swap. Equating this cost to zero, we find the value of V provided in Result 1.

The replicating strategy simplifies dramatically in the $\Delta \to 0$ limit. The dollar investment in each of the option portfolios expiring at times $\Delta, 2\Delta, \ldots, T-\Delta$ goes to zero at rate $O(\Delta^2)$. We must account, however, for the dynamically adjusted position in the underlying, indicated in rows beginning with a U. As shown in Table 3, this calls for a short position in the underlying asset of $2e^{-r(T-(j+1)\Delta)}S_{j\Delta}^2 e^{-\delta\Delta}/F_{j\Delta}^2$ in *dollar* terms at time $j\Delta$, i.e. a short position of $2e^{-r(T-(j+1)\Delta)}S_{j\Delta}e^{-\delta\Delta}/F_{j\Delta}^2$ units of the underlying. In the limit as $\Delta \to 0$, holding $j\Delta = t$ constant, this equates to a short position of $2e^{-r(T-t)}S_t/F_t^2$ units of the underlying asset at time t.

The static position in options expiring at time T, shown in the penultimate line of Table 2, does not disappear in the $\Delta \to 0$ limit. We can think of the option portfolio as a collection of calls of all strikes, as in (5). It is perhaps more natural, though, to think of the position as a collection of calls with strikes above F_T and puts with strikes below F_T, together with a long position in $2e^{-\delta(T-t)}/F_T$ units of the underlying asset—after continuous reinvestment of dividends—and a bond position. Combining this static long position in the underlying with the previously discussed dynamic position, the overall position at time t is long $2e^{-\delta(T-t)}/F_T - 2e^{-r(T-t)}S_t/F_t^2 = 2e^{-\delta(T-t)}(1 - S_t/F_t)/F_T$ units of the asset and long out-of-the-money-forward calls and puts, all financed by borrowing. Initially, therefore, the direct position in the underlying asset is $2(1 - S_0/F_0)/F_T = 0$ at time 0; subsequently, if the underlying's price at time t happens to exceed $F_t = S_0 e^{(r-\delta)t}$, then the replicating portfolio is short the underlying in order to offset the effects of increasing delta as calls go in-the-money and puts go increasingly out-of-the-money. In the other direction, if the underlying asset's price declines, then the delta-hedge requires buying the underlying to offset the negative delta resulting from puts going in-the-money and calls going out-of-the-money.

Result 3 (Hedging). *In the $\Delta \to 0$ limit, the payoff on a simple variance swap can be replicated by holding a portfolio, financed by borrowing, of*

(i) a static position in $2/F_T^2$ puts expiring at time T with strike K, for each $K \le F_T$;

(ii) a static position in $2/F_T^2$ calls expiring at time T with strike K, for each $K \ge F_T$; and

(iii) a dynamic position in $2e^{-\delta(T-t)}(1 - S_t/F_t)/F_T$ units of the underlying asset at time t.

These results continue to hold if any dividend payouts that occur are completely unanticipated at time 0. Consider an extreme case in which the simple variance swap is priced and hedged, at time zero, as though $\delta = 0$; but immediately after inception of the trade, at time $t = \Delta$, the underlying asset is suddenly liquidated via an extraordinary dividend, causing its (ex-dividend) price to equal 0 from time Δ onwards. The payout that must be made by the counterparty who is short variance is given by equation (1): in this extreme example, it will equal 1. Meanwhile, the hedge portfolio given in the above result will generate a positive payoff due to the put options going in-the-money. (The dynamic position will have zero payoff: it was neither long nor short at time 0, and subsequently the asset's price never moved from zero.) Since $S_T = 0$, the total payoff will be

$$\frac{2}{F_T^2}\int_0^{F_T} \max\{0, K - S_T\}\, dK = \frac{2}{F_T^2}\int_0^{F_T} K\, dK = 1.$$

In other words, the strategy perfectly replicates the desired payoff. More generally, this will continue to be the case as long as any dividends paid are completely unanticipated at time 0: once the strike V is set and the replicating portfolio is in place, it does not matter why the price path moves around subsequently, whether due to the payment of unanticipated dividends or not.

This logic does not apply if dividends are anticipated. In the appendix, I show how the definition of the payoff (1) must be modified if, instead of paying dividends at rate δS_t, the asset pays dividends whose sizes and timing are known at time 0. Once the definition is modified appropriately, the results above go through almost unchanged.

In the remainder of the main body of the paper I assume, for clarity, that the underlying asset does not pay dividends, $\delta = 0$, since this is the case considered in the related literature.

2 Variance swaps, VIX, and SVIX

In contrast, a (standard) variance swap pays

$$\left(\log\frac{S_\Delta}{S_0}\right)^2 + \left(\log\frac{S_{2\Delta}}{S_\Delta}\right)^2 + \cdots + \left(\log\frac{S_T}{S_{T-\Delta}}\right)^2 - \widetilde{V}$$

at time T. (This definition is very natural in a world in which asset prices follow diffusions.) To be fairly priced, we must have

$$\widetilde{V} = \mathbb{E}^*\left[\left(\log\frac{S_\Delta}{S_0}\right)^2 + \left(\log\frac{S_{2\Delta}}{S_\Delta}\right)^2 + \cdots + \left(\log\frac{S_T}{S_{T-\Delta}}\right)^2\right], \tag{8}$$

where the asterisk on the expectation indicates that it is taken with respect to the risk-neutral measure. To price the variance swap, i.e. to compute the expectation on the right-hand side of (8), it is generally assumed that the asset's price is an Itô process. When this is the case, we have the following result in the $\Delta \to 0$ limit; it is due to Carr and Madan (1998) and Demeterfi, Derman, Kamal, and Zou (1999), building on an idea of Neuberger (1994). From now on, $\widetilde{V}$ will always refer to the variance swap strike in the limiting case $\Delta \to 0$.

Result 4 (Neuberger (1994); Carr and Madan (1998); Demeterfi et al. (1999))**.** *If the underlying asset's price follows an Itô process $dS_t = rS_t\,dt + \sigma_t S_t\,dZ_t$ under the risk-neutral measure, then the strike on a variance swap is*

$$\widetilde{V} = 2e^{rT}\left\{\int_0^{F_T} \frac{1}{K^2}\,\mathsf{put}_T(K)\,dK + \int_{F_T}^{\infty} \frac{1}{K^2}\,\mathsf{call}_T(K)\,dK\right\}, \tag{9}$$

which has the interpretation

$$\widetilde{V} = \mathbb{E}^*\left[\int_0^T \sigma_t^2\,dt\right]. \tag{10}$$

The variance swap can be hedged by holding a portfolio, financed by borrowing, of

(i) a static position in $2/K^2$ puts expiring at time T with strike K, for each $K \le F_T$;

(ii) a static position in $2/K^2$ calls expiring at time T with strike K, for each $K \ge F_T$; and

(iii) a dynamic position in $2(F_t/S_t - 1)/F_T$ units of the underlying asset at time t.

Sketch of proof. In the $\Delta \to 0$ limit, the expectation (8) converges to[5]

$$\widetilde{V} = \mathbb{E}^*\left[\int_0^T (d\log S_t)^2\right].$$

Given the Itô process assumption, $d\log S_t = (r - \frac{1}{2}\sigma_t^2)dt + \sigma_t\,dZ_t$ under the risk-neutral measure, by Itô's lemma, so $(d\log S_t)^2 = \sigma_t^2\,dt$, and

$$\begin{aligned}
\widetilde{V} &= \mathbb{E}^*\left[\int_0^T \sigma_t^2\,dt\right] \\
&= 2\,\mathbb{E}^*\left[\int_0^T \frac{1}{S_t}\,dS_t - \int_0^T d\log S_t\right] \\
&= 2rT - 2\,\mathbb{E}^*\log\frac{S_T}{S_0}.
\end{aligned} \tag{11}$$

[5]To be more formal about it, $\widetilde{V}$ converges, under technical conditions established by Jarrow et al. (2010), to $\mathbb{E}^*[\langle \log S\rangle_T]$, where $\langle \log S\rangle_t$ is the quadratic variation process of $\log S_t$.

This shows that the strike on a variance swap is determined by pricing a notional contract that pays, at time T, the logarithm of the underlying asset's simple return $R_T = S_T/S_0$. Using Breeden-Litzenberger (1978) logic, the price of this contract, P_{log}, can be computed explicitly in terms of the prices of European call and put options on the underlying asset:

$$P_{\mathsf{log}} \equiv e^{-rT}\,\mathbb{E}^* \log R_T = rTe^{-rT} - \int_0^{F_T} \frac{1}{K^2}\,\mathsf{put}_T(K)\,dK - \int_{F_T}^{\infty} \frac{1}{K^2}\,\mathsf{call}_T(K)\,dK. \tag{12}$$

For completeness, a derivation is provided in the Appendix. Substituting (12) back into (11), we have the result. □

Another recent innovation, the gamma swap, is closely related to a variance swap. At time T, a gamma swap pays

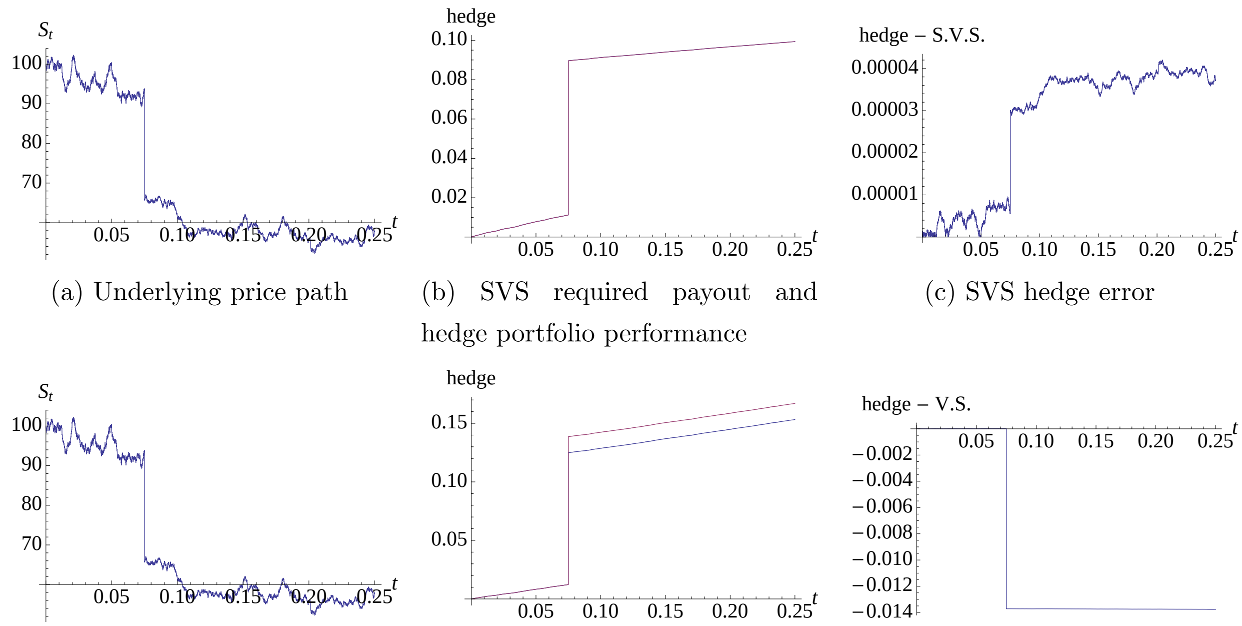
$$\frac{S_\Delta}{S_0}\left(\log\frac{S_\Delta}{S_0}\right)^2 + \frac{S_{2\Delta}}{S_0}\left(\log\frac{S_{2\Delta}}{S_\Delta}\right)^2 + \cdots + \frac{S_T}{S_0}\left(\log\frac{S_T}{S_{T-\Delta}}\right)^2 - V_\gamma\ .$$

Under the Itô process assumption, gamma swaps can be replicated in a similar way to variance swaps; see Lee (2010a). The replicating portfolio consists of static positions in out-of-the-money-forward calls and puts, in amounts inversely proportional to the strike (together with a dynamic delta-hedge, and money market positions for financing).

In contrast, the standard variance swap holds options in amounts inversely proportional to the *square* of the strike. Thus we have a sequence: variance swaps are hedged with a static portfolio of options with strikes K, held in amounts proportional to $1/K^2$; gamma swaps are hedged with a static portfolio of options with strikes K, held in amounts proportional to $1/K$; and simple variance swaps are hedged with a static portfolio of options with strikes K, held in amounts proportional to 1. Only for simple variance swaps, though, is replication possible in the presence of jumps.

Figure 1 makes this point graphically. It shows the required payoff and associated hedge portfolio payoff over a particular sample path for the underlying asset, for a 3-month variance swap and for a 3-month simple variance swap. Panels 1a, 1b, and 1c correspond to the simple variance swap, and panels 1d, 1e, and 1f to the variance swap. The underlying price follows the same path in each case. The middle panels compare the required payout on the simple variance swap and variance swap to the payout of the hedge portfolio in each case.[6] In

[6]For times t prior to expiry, I compute the required payout and hedge portfolio performance on the assumption that volatility goes to zero after time t, i.e. that the underlying asset price grows deterministically at the riskless rate between time t and time T.



(a) Underlying price path

(b) SVS required payout and hedge portfolio performance

(c) SVS hedge error

(d) Underlying price path

(e) VS required payout and hedge portfolio performance

(f) VS hedge error

Figure 1: A sample path of the required payout on a simple variance swap (SVS), and on a variance swap (VS), given the underlying price path shown in panels (a) and (d).

Panel 1b, only one line is visible: the simple variance swap is essentially perfectly hedged. In contrast, Panel 1e shows that the hedge portfolio (lower line) substantially underperforms the required payout on the variance swap (upper line) due to the downward jump in the price of the underlying. Panels 1c and 1f plot the difference between required payout and hedge portfolio performance in each case. In the case of the variance swap, the hedge error is enormous—on the order of 10% of the required payoff—while in the case of the simple variance swap, the corresponding error is three orders of magnitude smaller.[7]

Result 4 motivates the definition of the VIX index, which is calculated based on option prices using an annualized and discretized version of (9), and is generally interpreted as a measure of risk-neutral variance in the sense of (10). Working with the idealized version of VIX (i.e. not discretizing), we have

$$\mathrm{VIX}^2 \equiv \frac{2e^{rT}}{T}\left\{\int_0^{F_T} \frac{1}{K^2}\,\mathsf{put}_T(K)\,dK + \int_{F_T}^{\infty} \frac{1}{K^2}\,\mathsf{call}_T(K)\,dK\right\}.$$

[7]This example was generated using a discretization both in time—Δ greater than zero—and in the gap between strikes of options in the hedging portfolio. In the absence of such a discretization, the hedge error on a simple variance swap would be *exactly* zero, as shown in Result 3.

This is a *definition*, not a statement about pricing. Analogously, we can define an index, SVIX, based on the annualized strike of a simple variance swap. Based on (6), let SVIX be defined by

$$\text{SVIX}^2 \equiv \frac{e^{2rT}}{T} V(0,T) = \frac{2e^{rT}}{T \cdot S_0^2} \left\{ \int_0^{F_T} \mathsf{put}_T(K)\, dK + \int_{F_T}^{\infty} \mathsf{call}_T(K)\, dK \right\}. \tag{13}$$

This definition annualizes the strike on a simple variance swap, $V(0,T)$, and scales it by e^{2rT}. The scaling has two benefits: it makes SVIX more directly comparable to VIX, and it ensures that SVIX has a clean interpretation, as will be shown in Result 5.

Under the Itô process assumption, VIX^2 corresponds to the (annualized) strike on a variance swap, and has the interpretation (10). But this result leans heavily on the Itô process assumption: if there are jumps, the correctly priced strike $\widetilde{V}$ will not be given by (9); the replicating portfolio implied by the above analysis will not replicate the variance swap payoff, as Figure 1 shows; and neither $\widetilde{V}$ nor VIX^2 has the interpretation (10).[8] The next result shows what VIX *does* measure, and contrasts this with the much simpler interpretation of SVIX.

Result 5 (The interpretation of VIX and SVIX)**.** *Whether or not there are jumps, VIX measures the risk-neutral entropy of the simple return:*

$$VIX^2 = \frac{2}{T}\, \mathrm{L}^*(R_T), \tag{14}$$

where the entropy $\mathrm{L}^*(X) \equiv \log \mathbb{E}^* X - \mathbb{E}^* \log X$ *for positive random variables* X*. If the simple return* R_T *is lognormal, then*

$$VIX^2 = \frac{1}{T} \operatorname{var}^* \log R_T \approx \frac{1}{T} \operatorname{var}^* R_T, \tag{15}$$

where the approximation is accurate over short time horizons. But, in general, with jumps and/or time-varying volatility, VIX depends on all of the (annualized, risk-neutral) cumulants of log returns,

$$VIX^2 = 2 \sum_{n=2}^{\infty} \frac{\kappa_n^*}{n!} = \kappa_2^* + \frac{\kappa_3^*}{3} + \frac{\kappa_4^*}{12} + \frac{\kappa_5^*}{60} + \cdots, \tag{16}$$

where $\kappa_n^* \equiv \frac{1}{T}\widetilde{\kappa}_n^*$*, and* $\widetilde{\kappa}_n^*$ *is the nth cumulant of* $\log R_T$.[9]

[8] Up to a third-order approximation, Carr and Lee (2009) show that the fair variance swap strike is higher than that given in Result 4 if there are jumps and risk-neutral returns are negatively skewed.

[9] So, for example, $\widetilde{\kappa}_1^* = \mathbb{E}^* \log R_T$; $\widetilde{\kappa}_2^* = \operatorname{var}^* \log R_T$; $\widetilde{\kappa}_3^*$ is the skewness of $\log R_T$ multiplied by $(\widetilde{\kappa}_2^*)^{3/2}$; $\widetilde{\kappa}_4^*$ is the excess kurtosis, multiplied by $(\widetilde{\kappa}_2^*)^2$; and so on.

In contrast, SVIX measures the risk-neutral variance of the simple return:

$$SVIX^2 = \frac{1}{T} \operatorname{var}^* R_T. \tag{17}$$

Proof. Equation (14) follows from the definition of VIX^2 and (12), together with the fact that $\mathbb{E}^* R_T = e^{rT}$.

Equation (15) follows from (14) because $\log E^* R_T = \mathbb{E}^* \log R_T + \frac{1}{2} \operatorname{var}^* \log R_T$ if R_T is lognormal. For the approximation, write $\mu = \mathbb{E}^* \log R_T$ and $\sigma^2 = \operatorname{var}^* \log R_T$; over short time horizons, $\operatorname{var}^* R_T = e^{2\mu}\left(e^{2\sigma^2} - e^{\sigma^2}\right) = e^{2\mu}\sigma^2 + O(\sigma^4) \approx \sigma^2 = \operatorname{var}^* \log R_T$.

For the general result (16), we introduce the function $\boldsymbol{\kappa}^*(\theta) = \log \mathbb{E}^* \left[e^{\theta \cdot \log R_T}\right]$. This function can be expanded as a power series in θ,

$$\boldsymbol{\kappa}^*(\theta) = \sum_{n=1}^{\infty} \frac{\widetilde{\kappa}_n^* \theta^n}{n!},$$

where $\widetilde{\kappa}_n^*$ is the nth risk-neutral cumulant of $\log R_T$. The definition of entropy implies that $\mathrm{L}^*(R_T) = \boldsymbol{\kappa}^*(1) - \boldsymbol{\kappa}^{*\prime}(0)$, from which (16) follows after annualizing the cumulants: $\kappa_n^* \equiv \frac{1}{T}\widetilde{\kappa}_n^*$. For Normally distributed random variables, all cumulants above the variance are zero, so skewness and excess kurtosis (and so on) drop out in the lognormal case.

Finally, using asterisks to indicate variances and expectations with respect to the risk-neutral measure, we have

$$\operatorname{var}^* R_T = \mathbb{E}^* \left[\left(\frac{S_T}{S_0}\right)^2\right] - \left[\mathbb{E}^* \left(\frac{S_T}{S_0}\right)\right]^2 = \frac{e^{rT}\Pi(T)}{S_0^2} - e^{2rT}.$$

From (3), this implies

$$\operatorname{var}^* R_T = \frac{2e^{rT}}{S_0^2} \left\{ \int_0^{F_T} \mathsf{put}_T(K)\, dK + \int_{F_T}^{\infty} \mathsf{call}_T(K)\, dK \right\},$$

from which (17) follows. □

No part of Result 5 requires the Itô process assumption, because the derivation of equation (12) only depends on the static Breeden-Litzenberger logic. The entropy operator $\mathrm{L}(\cdot)$ provides a measure of the variability of a positive random variable. Like variance, it is nonnegative by Jensen's inequality, and like variance it measures variability by the extent to which a concave function of an expectation of a random variable exceeds an expectation of a concave function of a random variable. It has been applied in the finance literature by

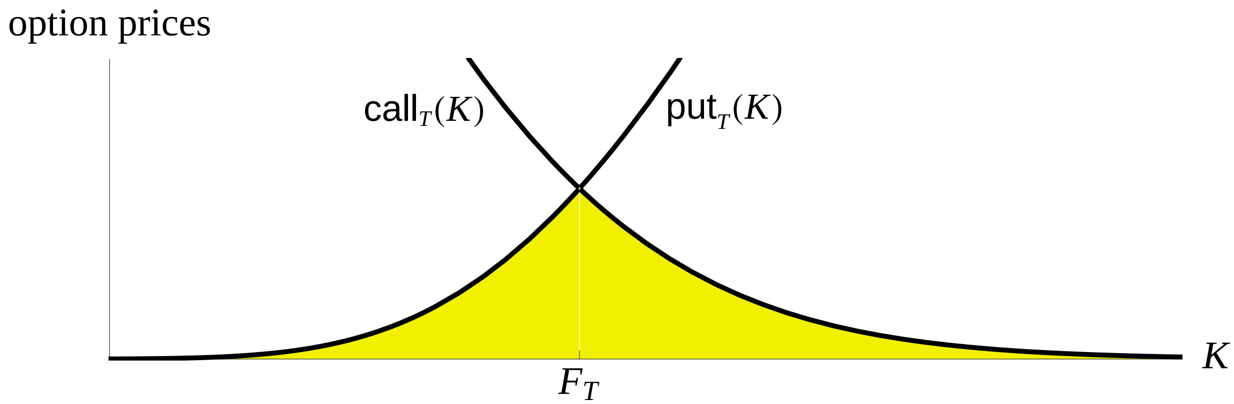


Figure 2: If the prices of call and put options expiring at time T are as shown, then the annualized risk-neutral variance of the underlying asset's simple return equals the shaded area under the curves multiplied by $2e^{rT}/(TS_0^2)$.

Alvarez and Jermann (2005), who refer to it as Theil's (1967) second entropy measure, and by Backus, Chernov and Martin (2010).[10]

Equation (17) has a nice graphical implication that is illustrated in Figure 2, which shows how to calculate the risk-neutral variance of the underlying asset's simple return to time T, given the prices of call and put options of all strikes expiring at time T. Calls and puts have equal value when the strike equals the forward price, so the two lines intersect at $K = F_T$. The annualized risk-neutral variance equals the shaded area under the two curves multiplied by $2e^{rT}/(TS_0^2)$. SVIX is the square root of this quantity, so measures risk-neutral volatility.

Equation (16) implies that if returns are more negatively skewed then, all else equal, VIX will be lower. At first glance, this seems surprising: one might have expected that negative skewness would drive VIX higher. This logic, of course, is based on intuition about real-world, not risk-neutral, cumulants; to assess it, we need to introduce some economics to create a link between the real-world probabilities and risk-neutral probabilities. We can do so by assuming that there is a representative agent with utility function $u(\cdot)$ who is content to hold the market—assumed to be the asset underlying VIX and SVIX—from to 0 to time T. Such an agent chooses from the available menu of assets with stochastic returns $R_T^{(i)}$, $i = 1, 2, \ldots$, and arrives at the overall portfolio return R_T. In other words, he chooses

[10] Entropy is an overloaded term. I use it here because of the link to Theil (1967). Backus, Chernov and Martin (2010) refer to $\mathrm{L}(M)$ as the entropy of a stochastic discount factor M because, in a complete market, $\mathrm{L}(M)$ equals the relative entropy, in the information-theoretic sense, of risk-neutral probabilities with respect to real-world probabilities.

portfolio weights $\{w_i\}$ to solve the maximization problem

$$\max_{\{w_i\}} \mathbb{E}\, u\left(\sum_i w_i R_T^{(i)}\right) \quad \text{subject to} \quad \sum_i w_i = 1. \tag{18}$$

The first-order conditions for this problem imply that $u'(R_T)/\,\mathbb{E}\,(R_T u'(R_T))$ is a stochastic discount factor. I use this fact in the proofs of the next two results.

Result 6 (Interpretation of VIX, continued)**.** *If there is a representative agent with power utility and coefficient of relative risk aversion γ, then VIX can be expressed in terms of the cumulants of $\log R_T$ under the real-world probabilities, i.e. $\kappa_2 = \frac{1}{T}\operatorname{var}\log R_T$ and so on:*

$$VIX^2 = \sum_{n=2}^{\infty} \alpha_n \kappa_n, \tag{19}$$

where $\alpha_2 = 1$ and $\alpha_3 = (1-3\gamma)/3$, and in general $\alpha_n = 2\left[(1-\gamma)^n - (-\gamma)^n - n(-\gamma)^{n-1}\right]/n!$.

Proof. Let $\boldsymbol{\kappa}(\theta) = \log \mathbb{E}\left[e^{\theta\cdot\log R_T}\right] = \sum_{n=1}^{\infty} \frac{\widetilde{\kappa}_n \theta^n}{n!}$ (the generating function of the real-world cumulants $\widetilde{\kappa}_n$), and let $\boldsymbol{\kappa}^*(\theta)$ be the corresponding CGF calculated with respect to risk-neutral probabilities. In this notation, equation (14) becomes

$$\text{VIX}^2 = \frac{2}{T}\left[\boldsymbol{\kappa}^*(1) - \boldsymbol{\kappa}^{*\prime}(0)\right]. \tag{20}$$

Since the representative investor has power utility, the SDF is proportional to $R_T^{-\gamma}$, so for any time-T payoff X, $\mathbb{E}^*\, X = \mathbb{E}\left(X R_T^{-\gamma}\right)/\,\mathbb{E}\left(R_T^{-\gamma}\right)$. From this it follows, on setting $X = R_T^{\theta}$, that $\boldsymbol{\kappa}^*(\theta) = \boldsymbol{\kappa}(\theta-\gamma) - \boldsymbol{\kappa}(-\gamma)$. Equation (20) becomes

$$\text{VIX}^2 = \frac{2}{T}\left[\boldsymbol{\kappa}(1-\gamma) - \boldsymbol{\kappa}(-\gamma) - \boldsymbol{\kappa}'(-\gamma)\right],$$

which gives (19) after annualizing the cumulants by writing $\kappa_n \equiv \frac{1}{T}\widetilde{\kappa}_n$. □

To interpret this result, note that if $\gamma \geq 1$, then α_n is positive for even n and negative for odd n. Result 6 therefore confirms the intuition discussed above: although VIX is positively related to risk-neutral skewness, it is negatively related to real-world skewness, and to all odd higher cumulants. It also shows that VIX is sensitively dependent, in a complicated way, on all the higher cumulants. In contrast, the next result shows that SVIX has an interpretation that is simpler and more interesting. I write $R_{f,T} = e^{rT}$ for the riskless rate to time T. The result requires that *any one* of the following assumptions holds:

A1 There is a one-period representative investor who maximizes (18) above, whose relative risk aversion $\gamma(x) \equiv -xu''(x)/u'(x)$, which need not be constant, satisfies $\gamma(x) \geq 1$.

A2 There is an intertemporal representative investor with separable utility whose value function J, which depends on current wealth W_0, can be defined recursively as

$$J\big[W_0\big] = \max_{C_0,\{w_i\}} u(C_0) + \beta\, \mathbb{E}\, J\Big[(W_0 - C_0)\sum_i w_i R_T^{(i)}\Big] \quad \text{s.t.} \ \sum_i w_i = 1,$$

and whose relative risk aversion $\Gamma(x) \equiv -xJ''\left[x\right]/J'\left[x\right]$, which need not be constant, satisfies $\Gamma(x) \geq 1$.

A3 The SDF and market return are conditionally jointly lognormal, and the market's conditional Sharpe ratio is greater than its conditional volatility.

A4 There is an Epstein-Zin (1989) representative investor with constant consumption-wealth ratio and risk aversion $\gamma \geq 1$.

These alternatives cover a variety of models. Assumption A1 is perhaps conceptually the simplest, but Cochrane (2011) has argued that it is crucial to allow for intertemporal effects in such calculations. Assumption A2 handles this case, and makes clear that the coefficient of relative risk aversion should be computed with respect to wealth, not with respect to consumption. This is the conventional measure of aversion to the risk of pure wealth bets. Assumption A3 covers the conditionally lognormal models of Campbell and Cochrane (1999) and Bansal and Yaron (2004), for calibrations that are consistent with the empirical regularity that the conditional Sharpe ratio of the market is higher than its conditional volatility. Finally, Assumption A4, which does not require any restriction on the distribution of returns, covers Wachter's (2011) time-varying disaster risk model.[11]

Result 7 (Interpretation of SVIX, continued)**.** *If any one of assumptions A1–A4 holds, then SVIX provides a lower bound on the equity premium:*

$$\frac{R_{f,T}}{T}\,\mathbb{E}\left[R_T - R_{f,T}\right] \geq SVIX^2. \tag{21}$$

If there is a representative agent with log utility, $\gamma(x) \equiv 1$*, then* (21) *holds with equality.*

[11] Wachter's model is in continuous time; A4 applies to a discretization in which the investor consumes at times 0, T, ..., and so on.

Proof. If there is a representative investor with log utility then $1/R_T$ is an SDF, so that $\frac{1}{R_{f,T}}\,\mathbb{E}^*\,X = \mathbb{E}\,[X/R_T]$ for any tradable X. The log utility case therefore follows from (17).

To handle the remaining cases we must show, given (17), that

$$\text{var}^*(R_T) \leq R_{f,T}\,(\mathbb{E}\,R_T - R_{f,T})\,. \tag{22}$$

Since $\mathbb{E}^*\,R_T = R_{f,T}$, this is equivalent to showing that $\frac{1}{R_{f,T}}\,\mathbb{E}^*\,(R_T^2) \leq \mathbb{E}\,R_T$, or that $\mathbb{E}\,(M_T R_T^2) \leq \mathbb{E}\,R_T$. This is equivalent to $\mathbb{E}\,(M_T R_T)\,\mathbb{E}\,R_T + \text{cov}\,(M_T R_T, R_T) \leq \mathbb{E}\,R_T$, where M_T is an SDF. Since $\mathbb{E}\,M_T R_T = 1$, this in turn is equivalent to showing that $\text{cov}\,(M_T R_T, R_T) \leq 0$. To establish this claim under any of assumptions A1, A2, and A4, we will show that $M_T R_T$ is decreasing in R_T, from which the result follows. Assumption A3 will be handled separately.

If A1 holds then, as mentioned above, the SDF is proportional to $u'(R_T)$. This reduces the problem to showing that $R_T u'(R_T)$ is decreasing in R_T, which holds because its derivative is $u'(R_T) + R_T u''(R_T) = -u'(R_T)\,[\gamma(R_T) - 1] \leq 0$.

If A2 holds, the first-order conditions for the investor's problem imply that $\beta J'[W_T]/J'[W_0]$ is an SDF. At this point we could invoke the envelope theorem result that $J'[W_t] = u'(C_t)$ to arrive at the SDF $\beta u'(C_T)/u'(C_0)$. Here, however, it is more convenient to think in terms of the value function J. Since the representative investor holds the market by definition, in equilibrium the chosen portfolio weights w_i are such that $W_T = (W_0 - C_0)R_T$. That is, the SDF is proportional to $J'\,[(W_0 - C_0)R_T]$. We must therefore show that $R_T J'\,[(W_0 - C_0)R_T]$ is a decreasing function of R_T. This follows as before: its derivative with respect to R_T is $-J'\,[W_T]\,\{\Gamma(W_T) - 1\}$, which is less than or equal to zero because $\Gamma(x) \geq 1$.

If A4 holds, then the SDF is equal (up to constants of proportionality that are known at time 0) to $C_T^{-\vartheta/\psi} R_T^{\vartheta-1}$, where $\vartheta \equiv (1-\gamma)/(1-1/\psi)$, with γ denoting the investor's risk aversion and ψ denoting the elasticity of intertemporal substitution; and the representative investor's wealth is proportional to the return on the market. Together, these imply that

$$M_T R_T \propto C_T^{-\vartheta/\psi} R_T^{\vartheta} = \left(\frac{C_T}{W_T}\right)^{-\vartheta/\psi} W_T^{-\vartheta/\psi} R_T^{\vartheta} \propto R_T^{\vartheta-\vartheta/\psi} = R_T^{1-\gamma},$$

which gives the result.

If A3 holds, we take a different approach. Write $M_T = e^{-r_f + \sigma_M Z_M - \sigma_M^2/2}$ and $R_T = e^{\mu_R + \sigma_R Z_R - \sigma_R^2/2}$, where Z_M and Z_R are correlated standard Normal random variables. Define $\lambda = (\mu_R - r_f)/\sigma_R$ to be the Sharpe ratio conditional on time-0 information. Given the lognormality assumption, some algebra shows that $\mathbb{E}\,M_T R_T^2 \leq \mathbb{E}\,R_T$ if and only if $\lambda \geq \sigma_R$. $\square$

That is, SVIX provides a direct measure of the expected risk premium on the market under the true, not the risk-neutral, probability distribution. This is a natural—perhaps *the* natural—measure of risk.

The theoretical content of Result 7 lies in inequality (22), which provides a bound in the opposite direction from the Hansen-Jagannathan (1991) bound:

$$\frac{\operatorname{var}^* R_T}{R_{f,T}} \le \mathbb{E}\, R_T - R_{f,T} \le R_{f,T} \cdot \sigma(M) \cdot \sigma(R_T).$$

The left inequality is (22); the right inequality is the Hansen-Jagannathan bound in the constant riskless rate case, with M denoting the SDF to time T and $\sigma(\cdot)$ denoting standard deviations computed with respect to the real-world probabilities. To exploit the Hansen-Jagannathan bound, the equity premium is generally treated as observable; this implies a lower bound on the volatility of the SDF. The important feature of (22) is that $\operatorname{var}^* R_T$, which implies a lower bound on the equity premium, and hence also on the product of SDF volatility and market return volatility, can be observed directly via option prices (or, in principle, simple variance swaps). The lower bound is of particular interest because, as the next section will show, it was strikingly high during the crisis of 2008–9.

As a general strategy for estimating the equity premium, the above result is reminiscent of an approach taken by Merton (1980), based on the equation

$$\text{instantaneous risk premium} = \gamma\sigma^2 \tag{23}$$

where γ is a measure of aggregate risk aversion, and σ^2 is the instantaneous variance of the market return. (Merton refers to this as the "'Constant Preferences' Model #1".) This relationship can be justified in an environment in which there is a representative agent with power utility and coefficient of relative risk aversion γ.

There are some important differences between the two approaches. In particular, Merton assumes that the market's price follows a diffusion. In contrast, Result 7 makes no assumption about how prices evolve. Related to this, there is no distinction between risk-neutral and real-world (instantaneous) variance in a diffusion-based model: the two are identical, by Girsanov's theorem. Inequality (22) shows that once we move beyond diffusions, the appropriate generalization relates the *real-world* risk premium to the *risk-neutral* variance.

A second difference is that assumptions A1 and A2 do not require that there is a representative agent with *constant* relative risk aversion, only that the agent's risk aversion is never less than 1. This has the advantage of increased generality but the apparent disadvantage that, other than in the log utility case, Result 7 only provides a lower bound. It turns out, though, that this apparent disadvantage reflects a real phenomenon rather than a suboptimal result. One might imagine, for example, that in the CRRA case $\gamma(x) \equiv \gamma$, we could show that $\gamma \cdot \text{SVIX}^2 = \frac{R_{f,T}}{T}\,\mathbb{E}\,[R_T - R_{f,T}]$. It can be shown, though, that this fails for $\gamma \neq 1$. Similarly, one might hope to show that if $\gamma(x) \geq \gamma$, then $\gamma \cdot \text{SVIX}^2 \leq \frac{R_{f,T}}{T}\,\mathbb{E}\,[R_T - R_{f,T}]$. But this does not follow either; nor does the corresponding result with the two inequalities reversed. In every case, higher cumulants can conspire to invalidate the hoped-for conclusion.

Third, Merton implements (23) using realized historical volatility rather than by exploiting option price data, though he notes that volatility measures could be calculated "by 'inverting' the Black-Scholes option pricing formula".[12] However, Black-Scholes implied volatility would only provide the correct measure of σ if we really lived in a Black-Scholes (1973) world with lognormal prices and constant volatility. The results of this paper show how to compute the right measure of variance in a more general environment.

3 Applications of SVIX

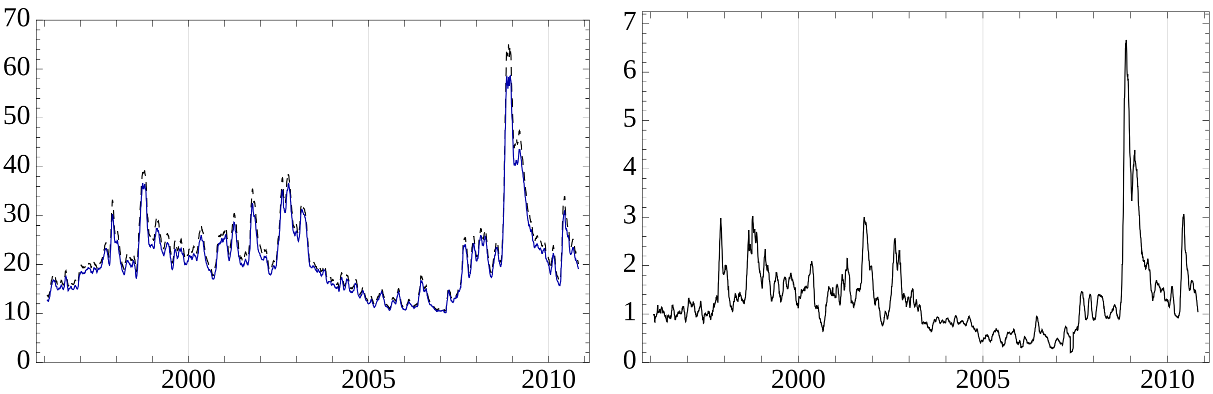
3.1 A lower bound on the equity premium

I construct the time series of SVIX using option price data supplied by *OptionMetrics*, following the methodology used to construct VIX.[13] Full details of the procedure are in the Appendix. The underlying asset is the S&P 500 index, and I compute the index for horizons of $T = 1$, 2, 3, 6, and 12 months.

Figures 3a and 3b plot SVIX in the 1 month case over 3,729 days from January 4, 1996, to October 29, 2010. For clarity, the figures show 20-day moving averages. Figure 3a shows the time series of VIX (dotted line) and SVIX (solid line) at each day's close. At the scale of the figure, it is hard to see any difference between the two, though VIX's sensitivity to higher cumulants is visible at some of the peaks. Figure 3b plots VIX minus SVIX. In theory, it is possible for VIX to be lower than SVIX. Indeed, this would (just) occur if returns R_T were lognormally distributed under the risk-neutral measure. Roughly speaking, Figure 3b captures the variation over time in the higher risk-neutral moments of returns. More precisely, Result 5 shows that the difference between VIX squared and SVIX squared is

[12] In principle—unfortunately index options were not traded at the time he wrote the paper.

[13] See *The CBOE Volatility Index – VIX*, available on the CBOE website.



(a) VIX and SVIX

(b) VIX minus SVIX

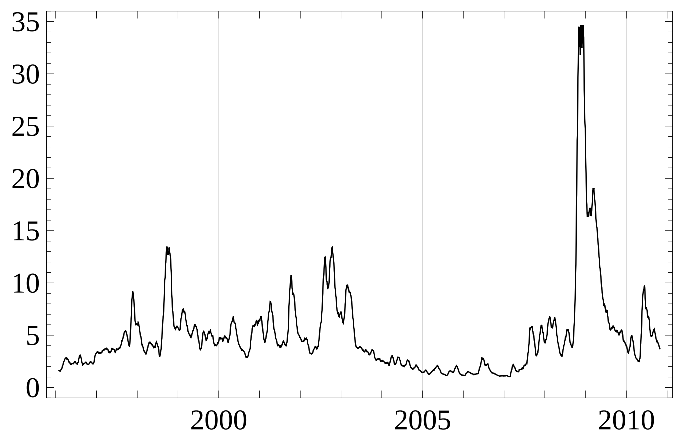
Figure 3: Left: Time series of closing prices of VIX (dotted line) and SVIX (solid line). Right: VIX minus SVIX.

$\frac{1}{T}(\mathrm{L}^*(R_T) - \mathrm{var}^* R_T)$. Unsurprisingly, therefore, the VIX-SVIX spread jumped in the recent crisis and at other times of market stress.

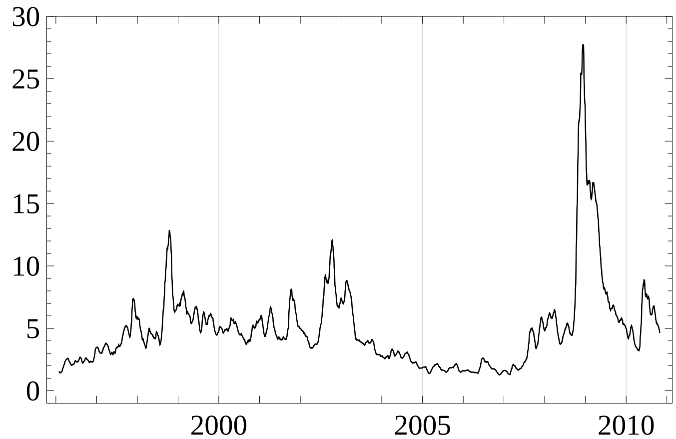
Continuing with the one-month horizon, Figure 4a plots the 20-day moving average of $\frac{1}{R_{f,T}}\mathrm{SVIX}^2$ measured in percentage points. As shown in Result 7, this can be interpreted as a lower bound on the annualized expected equity premium. The mean of this lower bound over the whole sample is 4.99%—a number close to typical estimates of the unconditional equity premium. If either Assumption A1 or A2 holds, this raises the possibility that the marginal stock market investor's relative risk aversion is closer to 1 than is commonly supposed in the literature. There is dramatic time-variation in the lower bound. If we split the sample into two, with December 31, 2007 as the dividing line, then the average lower bound on the equity premium is 4.09% in the early period and 8.78% in the later period. During the years 2004–2006, the average lower bound was only 1.86%; in contrast, during the recent crisis, the lower bound peaked at 34.7% on a 20-day moving average basis, and rose as high as 55.0% in the daily data.

Figures 4b and 4c repeat this exercise for the 3-month and 1-year horizons. The pattern is the same, though as one would expect the bound is less volatile as the horizon length increases. Nonetheless, even at the annual horizon, there is substantial variation, from a minimum of 1.22% to a maximum of 21.5% in the daily data.

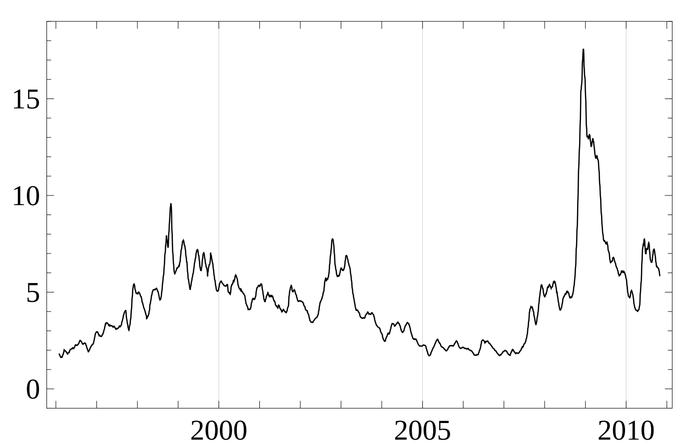
At all horizons, the equity premium hit peaks during the recent crisis, notably from late 2008 to early 2009 as the credit crisis gathered steam and the stock market fell, but



(a) 1 month



(b) 3 months



(c) 1 year

Figure 4: The lower bound on the annualized equity premium at different horizons (in %).

also around May 2010, coinciding with the beginning of the European sovereign debt crisis. Other visible peaks occur during the LTCM crisis in late 1998; during the days following September 11, 2001; and during a period in late 2002 when the stock market was hitting new lows following the end of the dotcom boom. Interestingly, the bound was also relatively high from late 1998 until the end of 1999; by contrast, forecasts based on market dividend- and earnings-price ratios incorrectly predicted an extraordinarily low or even negative equity premium during this period, as noted by Ang and Bekaert (2007) and Goyal and Welch (2008). Moreover, the out-of-sample issues emphasized by Goyal and Welch (2008) do not apply here, since no parameter estimation is required to generate the lower bounds. Finally, by construction, the lower bound can never be less than zero.

horizon	mean	s.d.	min	1%	10%	25%	50%	75%	90%	99%	max
1 mo	4.99	4.68	0.83	1.02	1.50	2.39	3.95	5.69	8.77	26.4	55.0
2 mo	4.98	4.06	1.01	1.20	1.61	2.52	4.15	5.89	8.43	24.6	46.1
3 mo	4.93	3.67	1.07	1.29	1.72	2.56	4.27	5.90	8.01	21.9	39.1
6 mo	4.84	3.02	1.30	1.53	1.92	2.66	4.38	5.97	7.48	17.2	29.0
1 yr	4.57	2.48	1.22	1.64	2.03	2.62	4.26	5.63	7.09	14.1	21.5

Table 1: Mean, standard deviation, and quantiles of annualized EP bound (in %).

Table 1 reports the mean, standard deviation, and various quantiles of the distribution of the lower bound in the daily data for horizons between 1 month and 1 year. It is worth emphasizing that although VIX is more positively skewed and has a higher kurtosis than SVIX, the quantity that enters the equity premium bound is SVIX *squared*, which in turn is more skewed and has a higher kurtosis than the VIX index.

3.2 Market-implied correlation

The strikes of simple variance swaps on the market and of simple variance swaps on the constituents of the market can be combined to supply a measure of market-implied (i.e. risk-neutral) correlation. Write σ_M^2 for the risk-neutral variance of the simple return on the market from time 0 to time T; write σ_i^2 for the risk-neutral variance of the simple return on stock i from time 0 to time T; write ρ_{ij} for the correlation between stocks i and j; and write

w_i for the market weights:

$$\sigma_M^2 = \text{var}^* \left(\sum_{i=1}^{N} w_i R_i \right) = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i \neq j} w_i w_j \rho_{ij} \sigma_i \sigma_j \, . \tag{24}$$

For the purposes of backing out a correlation measure, we need the variances of *simple returns*, not log returns: because the quantity $\text{var}^* \log \sum w_i R_i$ does not decompose nicely into a sum of individual variances $\text{var}^* \log R_i$, there is no analogue of (24) for log returns. Thus simple variance swaps would provide the natural way to measure implied correlation even in a world in which jumps did not occur.

Based on (24), we can define the implied correlation measure

$$\widehat{\rho} = \frac{\sigma_M^2 - \sum w_i^2 \sigma_i^2}{\sum_{i \neq j} w_i w_j \sigma_i \sigma_j} \, .$$

The implied correlation $\widehat{\rho}$ can then be calculated directly from the strike on a simple variance swap on the market (which reveals σ_M^2, via equations (6) and (17)) and from the strikes on simple variance swaps on the index constituents (which reveal $\{\sigma_i^2\}_{i=1,\ldots,N}$), together with the observable market weights w_i.

Computing a measure of market-implied correlation from (standard) variance swaps is considerably more challenging. Even if prices were continuous, the strike on a variance swap would not reveal risk-neutral variance, but the risk-neutral expectation of integrated instantaneous variance, as in equation (10). And even if the index constituents had log-normal simple returns $R_T^{(i)}$, so that variance swaps on the underlying assets revealed the risk-neutral variance of log returns, $\text{var}^* \log R_T^{(i)}$, as in (15), the market return itself—a sum of lognormals—would not be lognormal. In the presence of jumps, these problems are even more acute. Most obviously, even setting aside the difficulties in pricing variance swaps in the presence of jumps and in using the resulting prices to compute a correlation measure, it is critical that the single-name market should not evaporate at times of stress, since the variances of single names are needed to compute correlation.

4 Conclusion

The market turmoil of 2008 and 2009 should have been the volatility derivatives market's moment in the sun. Unfortunately, at the moment when the ability to hedge and speculate in volatility and variance would have been particularly valuable, these markets dried up.

Why was this? The critical assumption that underpins the theory of pricing and hedging of variance swaps is that prices follow diffusions, and hence cannot jump. The dependence on this assumption means that variance swaps are hardest to price at times when they are needed most. This problem is particularly severe for variance swaps on single names: Carr and Lee (2009) observe that "the contractual payoffs that appear in thousands of term sheets become literally infinite if the underlying ever closes at zero." As a workaround, it is now the market convention to cap the payoffs on variance swaps, though doing so complicates the pricing and destroys the clean interpretation of the contract.

This paper has proposed a financial contract, a simple variance swap, that is closely related to a traditional variance swap. Unlike variance swaps, however, simple variance swaps can be priced and hedged even in the presence of jumps. The weighting by forward price in the definition (1) leads to an important simplification of the hedging strategy that is critical to make the contract potentially tradable in practice. Simple variance swaps have a natural interpretation: whether or not there are jumps, they reveal the (risk-neutral) variance of the simple return on the underlying asset, so permit hedging and speculation by market participants with views on this quantity.

The contrast between simple variance swaps and variance swaps can be seen most directly by examining their respective hedging portfolios. The hedge portfolio for a simple variance swap holds equal amounts of options of all different strikes, while variance swaps require increasingly large positions in puts with increasingly low strikes. This makes explicit the dependence of variance swaps on extreme events. Even if one is prepared to assert that there are no jumps (as is required to legitimize the theory of variance swap and gamma swap pricing), variance swaps are harder to hedge than simple variance swaps, since they load more strongly on deep-out-of-the-money puts.

As a result of these unfortunate properties of variance swaps, gamma swaps have recently started trading. While the hedge portfolio for variance swaps holds portfolios of options with strikes K in amounts proportional to $1/K^2$, the hedge for a gamma swap holds portfolios of options with strikes K in amounts proportional to $1/K$. However, as with variance swaps, this only holds if prices cannot jump. In the sense that the hedge portfolio on a simple variance swap holds options with strikes K in amounts proportional to 1, simple variance swaps are the answer to the question: 'What is the next member of the sequence "variance swap, gamma swap, ... "?' But simple variance swaps are distinguished from the other two by the fact that they can be priced and hedged in the presence of jumps.

The paper has also proposed two applications of simple variance swaps. They have a natural application to measuring implied correlation. They can also be used to derive a forward-looking lower bound on the expected equity premium (computed with respect to real-world, not risk-neutral, probabilities). In the data, the resulting lower bounds, which can be seen in Figure 4, are striking. The mean lower bound over the full sample—at 4.99% in annualized terms for returns over a one-month horizon and 4.57% for a one-year horizon—is close to the long-run average realized equity premium. But the lower bound is extraordinarily volatile: at the peak of the recent crisis, in November 2008, the lower bound on the one-year equity premium rose to 21.5%, while the lower bound on the annualized one-month equity premium climbed to 55.0%.

5 Bibliography

Alvarez, F., and U. J. Jermann (2005), "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73:6:1977–2016.

Ang, A., and G. Bekaert (2007), "Stock Return Predictability: Is It There?", *Review of Financial Studies*, 20:3:651–707.

Backus, D., M. Chernov, and I. Martin (2010), "Disasters Implied by Equity Index Options", *Journal of Finance*, forthcoming.

Bansal, R. and A. Yaron (2004), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59:4:1481–1509.

Black, F., and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81:637–659.

Breeden, D. T., and R. H. Litzenberger (1978), "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51:4:621–651.

Campbell, J. Y. and J. H. Cochrane (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107:2:205–251.

Carr, P., and A. Corso (2001), "Covariance Contracting for Commodities," *EPRM* April 2001.

Carr, P., and R. Lee (2009), "Volatility Derivatives," *Annual Review of Financial Economics*, 1:1–21.

Carr, P., and D. Madan (1998), "Towards a Theory of Volatility Trading," in R. Jarrow, ed., *Volatility: New Estimation Techniques for Pricing Derivatives*, London: Risk Books, pp. 417–427.

Cochrane, J. H. (2011), "Discount Rates," *Journal of Finance*, forthcoming.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou (1999), "More Than You Ever Wanted to Know about Volatility Swaps," Goldman Sachs Quantitative Strategies Research Notes.

Epstein, L., and S. Zin (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57:937–969.

Goyal, A., and I. Welch (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21:4:1455–1508.

Hansen, L. P. and R. Jagannathan (1991), "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99:2:225–262.

Jarrow, R. A., P. Protter, M. Larsson, and Y. Kchia (2010), "Variance and Volatility Swaps: Bubbles and Fundamental Prices," Cornell University, Johnson School Research Paper.

Lee, R. (2010a), "Gamma Swap," in Rama Cont, ed., *Encyclopedia of Quantitative Finance*, Wiley.

Lee, R. (2010b), "Weighted Variance Swap," in Rama Cont, ed., *Encyclopedia of Quantitative Finance*, Wiley.

Merton, R. C. (1980), "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, 8:323–361.

Neuberger, A. (1990), "Volatility Trading," working paper, London Business School.

Neuberger, A. (1994), "The Log Contract," *Journal of Portfolio Management*, 20:2:74–80.

Summers, L. (1985), "On Economics and Finance," *Journal of Finance*, 40:3:633–635.

Theil, H. (1967), *Economics and Information Theory*, Amsterdam: North-Holland.

Wachter, J. (2011), "Can time-varying risk of rare disasters explain aggregate stock market volatility?", working paper, Wharton School of Business.

A Appendix

A.1 Using the $\Delta \to 0$ approximation if $\Delta > 0$

This section shows that the expression (6) is extremely accurate so long as Δ is reasonably small by deriving an analytic upper bound on the magnitude of the approximation error, and then showing that the upper bound is tiny in practice.

Result 8. *For* Δ *greater than zero,* $V(\Delta, T)$ *is very well approximated by* $\frac{e^{rT}P(T)}{F_{T-\Delta}^2}$*:*

$$\left| V(\Delta,T) - \frac{e^{rT}P(T)}{F_{T-\Delta}^2} \right| \le \frac{T}{\Delta}\left(e^{|r-\delta|\Delta} - 1\right)^2 \cdot \left(1 + \frac{e^{rT}P(T)}{F_{T-\Delta}^2}\right). \tag{25}$$

Proof. Result 2 showed that

$$V(0,t) = \lim_{\Delta\to 0} V(\Delta,t) = \lim_{\Delta\to 0} \mathbb{E}^* \left(\sum_{i=1}^{t/\Delta} \left[\frac{S_{i\Delta} - S_{(i-1)\Delta}}{F_{(i-1)\Delta}} \right]^2 \right) = \frac{e^{rt}P(t)}{F_t^2}.$$

Therefore, whenever t_1 and t_2 are both multiples of Δ with $t_1 \le t_2$, we have

$$\frac{e^{rt_1}P_{t_1}}{F_{t_1}^2} = \lim_{\Delta\to 0} \mathbb{E}^* \left(\sum_{i=1}^{t_1/\Delta} \left[\frac{S_{i\Delta} - S_{(i-1)\Delta}}{F_{(i-1)\Delta}} \right]^2 \right) \le \lim_{\Delta\to 0} \mathbb{E}^* \left(\sum_{i=1}^{t_2/\Delta} \left[\frac{S_{i\Delta} - S_{(i-1)\Delta}}{F_{(i-1)\Delta}} \right]^2 \right) = \frac{e^{rt_2}P(t_2)}{F_{t_2}^2},$$

so that $e^{rt}P(t)/F_t^2$ is increasing in t. It follows that

$$\begin{aligned}
\left| V(\Delta,T) - \frac{e^{rT}P(T)}{F_{T-\Delta}^2} \right| &= \sum_{j=1}^{T/\Delta - 1} \left(e^{(r-\delta)\Delta} - 1\right)^2 \frac{e^{rj\Delta}P(j\Delta)}{F_{j\Delta}^2} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2 \\
&\le \left(\frac{T}{\Delta} - 1\right)\left(e^{(r-\delta)\Delta} - 1\right)^2 \frac{e^{rT}P(T)}{F_T^2} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2 \\
&= \left(\frac{T}{\Delta} - 1\right)\left(e^{-(r-\delta)\Delta} - 1\right)^2 \frac{e^{rT}P(T)}{F_{T-\Delta}^2} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2 \\
&\le \frac{T}{\Delta}\left(e^{|r-\delta|\Delta} - 1\right)^2 \cdot \left(1 + \frac{e^{rT}P(T)}{F_{T-\Delta}^2}\right),
\end{aligned}$$

as required. □

The term $\frac{T}{\Delta}\left(e^{|r-\delta|\Delta} - 1\right)^2$ is *tiny* for sensible parameter values. For comparison, Table 1 shows that $V(\Delta,T)$ and $e^{rT}P(T)/F_{T-\Delta}^2$ have means of about $0.05T$. (Note that Table 1 reports quantities scaled by $1/T$.) So the upper bound on the *relative error* implied by the above result, i.e. the percentage inaccuracy from using expression (6) when $\Delta > 0$, is

$$\frac{\frac{T}{\Delta}\left(e^{|r-\delta|\Delta} - 1\right)^2 \cdot \left(1 + \frac{e^{rT}P(T)}{F_{T-\Delta}^2}\right)}{\frac{e^{rT}P(T)}{F_{T-\Delta}^2}} \sim \frac{20}{\Delta}\left(e^{|r-\delta|\Delta} - 1\right)^2 (1 + 0.05T).$$

The size of T makes essentially no difference to the calculations, but for concreteness consider a 1-year simple variance swap, $T = 1$. If $r - \delta = 3\%$ then the term on the right-hand side is 0.008% if the simple variance swap is based on daily price changes ($\Delta = 1/252$), and 0.04% if the simple variance swap is based on weekly price changes ($\Delta = 1/52$). Even if $\Delta = 0.5$, the percentage error is still less than 1%.

A.2 A derivation of (12)

Using the result of Breeden and Litzenberger (1978), we have

$$P_{\mathsf{log}} = \int_0^\infty \log \frac{K}{S_0}\, \mathsf{call}''_T(K)\, dK.$$

Evaluating this integral is a straightforward exercise in integration by parts, though we do need one small trick right at the beginning, splitting the range of integration into two parts and using the observation that $\mathsf{put}''_T(K) \equiv \mathsf{call}''_T(K)$, which follows from put-call parity; and, half-way through, to use the fact that $\mathsf{put}'_T(K) - \mathsf{call}'_T(K) = e^{-rT}$, which again follows from put-call parity.

$$\begin{aligned}
P_{\mathsf{log}} &= \int_0^{F_T} \log \frac{K}{S_0}\, \mathsf{put}''_T(K)\, dK + \int_{F_T}^\infty \log \frac{K}{S_0}\, \mathsf{call}''_T(K)\, dK \\
&= \left[\log \frac{K}{S_0} \cdot \mathsf{put}'_T(K)\right]_0^{F_T} - \int_0^{F_T} \frac{1}{K}\, \mathsf{put}'_T(K)\, dK + \left[\log \frac{K}{S_0} \cdot \mathsf{call}'_T(K)\right]_{F_T}^\infty - \int_{F_T}^\infty \frac{1}{K}\, \mathsf{call}'_T(K)\, dK \\
&= rTe^{-rT} - \int_0^{F_T} \frac{1}{K}\, \mathsf{put}'_T(K)\, dK - \int_{F_T}^\infty \frac{1}{K}\, \mathsf{call}'_T(K)\, dK \\
&= rTe^{-rT} + \left[\frac{-\,\mathsf{put}_T(K)}{K}\right]_0^{F_T} - \int_0^{F_T} \frac{1}{K^2}\, \mathsf{put}_T(K)\, dK + \left[\frac{-\,\mathsf{call}_T(K)}{K}\right]_{F_T}^\infty - \int_{F_T}^\infty \frac{1}{K^2}\, \mathsf{call}_T(K)\, dK \\
&= rTe^{-rT} - \int_0^{F_T} \frac{1}{K^2}\, \mathsf{put}_T(K)\, dK - \int_{F_T}^\infty \frac{1}{K^2}\, \mathsf{call}_T(K)\, dK.
\end{aligned}$$

A.3 Construction of SVIX

The data are from *OptionMetrics*, runnning from January 4, 1996, to October 29, 2010; they include the closing price of the S&P 500 index, and the expiration date, strike price, highest closing bid and lowest closing ask of all call and put options with fewer than 550 days to expiry. I clean the data in several ways. First, I delete all replicated entries (of which there are more than 500,000). Second, I delete all calls with strikes below the spot and puts with strikes above the spot price. Third, I delete all options with a highest closing bid of zero. Finally, I delete all Quarterly options, which tend to be less liquid than regular S&P 500 index options and to have a smaller range of strikes. Having done so, I am left with 985,219 option-day datapoints. I compute mid-market option prices by averaging the highest closing bid and lowest closing ask, and using the resulting prices to compute SVIX via a discretization of equation (13).

This procedure implicitly approximates the forward price F_T by the spot price S_0. To understand the size of the error that this introduces, we can use put-call parity to rewrite (13) as

$$\text{SVIX}^2 = \frac{2e^{rT}}{T \cdot S_0^2} \left\{ \int_0^{S_0} \mathsf{put}_T(K)\, dK + \int_{S_0}^{\infty} \mathsf{call}_T(K)\, dK \right\} - \frac{1}{T} \left(e^{rT} - 1 \right)^2 .$$

By approximating the forward price by the spot price in (13), I am neglecting the final term on the right-hand side. This term is very small: with $T = 0.25$, $r = 0.03$, it equals 0.0002. On any given day, I compute SVIX for a range of time horizons depending on the particular expiration dates of options traded on that day, with the constraint that the shortest time to expiry is never allowed to be less than 7 days; this is the same procedure that the CBOE follows. I then calculate the implied SVIX for $T = 30$, 60, 90, 180, and 360 days by linear interpolation. Occasionally, extrapolation is necessary, for example when the nearest-term option's time-to-maturity first dips below 7 days, requiring me to use the two expiry dates further out; this is also the case with the CBOE's calculation of VIX.

asset	0	Δ	2Δ	$\dots$	$T-\Delta$	T
B	$-e^{-rT}$			$\dots$		$\frac{S_0^2}{S_0^2}$
U	$2e^{-r(T-\Delta)}e^{-\delta\Delta}$	$-2e^{-r(T-\Delta)}\frac{S_\Delta}{S_0}$		$\dots$		
B		$2e^{-r(T-\Delta)}\frac{S_\Delta}{S_0}$		$\dots$		$-2\frac{S_0 S_\Delta}{S_0^2}$
Δ	$-\frac{\left(e^{(r-\delta)\Delta}-1\right)^2\Pi_\Delta}{e^{r(T-\Delta)}F_\Delta^2}$	$\frac{\left(e^{(r-\delta)\Delta}-1\right)^2 S_\Delta^2}{e^{r(T-\Delta)}F_\Delta^2}$		$\dots$		
B		$-e^{-r(T-\Delta)}\left(\frac{S_\Delta^2}{S_0^2}+\frac{S_\Delta^2}{F_\Delta^2}\right)$		$\dots$		$\frac{S_\Delta^2}{S_0^2}+\frac{S_\Delta^2}{F_\Delta^2}$
U		$\frac{2e^{-r(T-2\Delta)}S_\Delta^2}{F_\Delta^2}e^{-\delta\Delta}$	$\frac{-2e^{-r(T-2\Delta)}S_\Delta S_{2\Delta}}{F_\Delta^2}$	$\dots$		
B			$\frac{2e^{-r(T-2\Delta)}S_\Delta S_{2\Delta}}{F_\Delta^2}$	$\dots$		$\frac{-2S_\Delta S_{2\Delta}}{F_\Delta^2}$
$\vdots$	$\vdots$			$\dots$		$\vdots$
$T-\Delta$	$-\frac{\left(e^{(r-\delta)\Delta}-1\right)^2\Pi_{T-\Delta}}{e^{r\Delta}F_{T-\Delta}^2}$			$\dots$	$\frac{\left(e^{(r-\delta)\Delta}-1\right)^2 S_{T-\Delta}^2}{e^{r\Delta}F_{T-\Delta}^2}$	
B				$\dots$	$-e^{-r\Delta}\left(\frac{S_{T-\Delta}^2}{F_{T-2\Delta}^2}+\frac{S_{T-\Delta}^2}{F_{T-\Delta}^2}\right)$	$\frac{S_{T-\Delta}^2}{F_{T-2\Delta}^2}+\frac{S_{T-\Delta}^2}{F_{T-\Delta}^2}$
U				$\dots$	$\frac{2S_{T-\Delta}^2}{F_{T-\Delta}^2}e^{-\delta\Delta}$	$\frac{-2S_{T-\Delta}S_T}{F_{T-\Delta}^2}$
T	$-\frac{\Pi_T}{F_{T-\Delta}^2}$			$\dots$		$\frac{S_T^2}{F_{T-\Delta}^2}$
B	Ve^{-rT}			$\dots$		$-V$

Table 2: Replicating the simple variance swap. In the left column, B indicates dollar positions in the bond, U indicates dollar positions in the underlying with dividends continuously reinvested, and $j\Delta$, for $j = 1, 2, \dots, T/\Delta$, indicates a position in the portfolio of options expiring at time $j\Delta$ that replicates the payoff $S_{j\Delta}^2$, whose price at time 0 is $\Pi_{j\Delta}$.

asset	0	$j\Delta$	$(j+1)\Delta$	T
$j\Delta$	$-\frac{\left(e^{(r-\delta)\Delta}-1\right)^2\Pi_{j\Delta}}{e^{r(T-j\Delta)}F^2_{j\Delta}}$	$\frac{\left(e^{(r-\delta)\Delta}-1\right)^2 S^2_{j\Delta}}{e^{r(T-j\Delta)}F^2_{j\Delta}}$		
B		$-e^{-r(T-j\Delta)}\left(\frac{S^2_{j\Delta}}{F^2_{(j-1)\Delta}}+\frac{S^2_{j\Delta}}{F^2_{j\Delta}}\right)$		$\frac{S^2_{j\Delta}}{F^2_{(j-1)\Delta}}+\frac{S^2_{j\Delta}}{F^2_{j\Delta}}$
U		$\frac{2S^2_{j\Delta}e^{-\delta\Delta}}{e^{r(T-(j+1)\Delta)}F^2_{j\Delta}}$	$\frac{-2S_{j\Delta}S_{(j+1)\Delta}}{e^{r(T-(j+1)\Delta)}F^2_{j\Delta}}$	
B			$\frac{2S_{j\Delta}S_{(j+1)\Delta}}{e^{r(T-(j+1)\Delta)}F^2_{j\Delta}}$	$\frac{-2S_{j\Delta}S_{(j+1)\Delta}}{F^2_{j\Delta}}$

Table 3: Replicating the simple variance swap. The generic position in options of intermediate maturity, together with the associated trades required after expiry. In the left column, B indicates a position in the bond, U indicates a position in the underlying with dividends continuously reinvested, and $j\Delta$ indicates a position in options expiring at $j\Delta$.

A.4 The case in which the underlying pays known dividends

For simplicity, consider the case in which the asset pays a single dividend $D_{k\Delta}$ at time $k\Delta$ for some k, and no dividends at any other time up to and including the expiry date, T. The price of a portfolio whose payoff is S_i^2 at time i continues to equal $\Pi(i)$, given by equation 5.

In this section, it will be important to distinguish between F_t, the forward price of the dividend-paying asset to time t, and $\widetilde{F}_t \equiv S_0 e^{rt}$, the appropriate normalization for the definition of a simple variance swap in this case. A standard no-arbitrage argument implies that the forward price is given by

$$F_t = \begin{cases} S_0 e^{rt} & \text{if } t < k\Delta \\ S_0 e^{rt} - D_{k\Delta} e^{r(t-k\Delta)} & \text{if } t \geq k\Delta \end{cases}$$

so F_t and $\widetilde{F}_t$ coincide for times t before the payment of the dividend, but differ thereafter. It turns out that $\widetilde{F}_t$ is the appropriate normalization so that the intermediate option positions are negligibly small, as was the case in the main text.

The definition of the payoff on the simple variance swap must be modified to allow for the presence of the dividend. At time T, the counterparties to the simple variance swap now exchange V for

$$\left(\frac{S_\Delta - S_0}{S_0}\right)^2 + \cdots + \left(\frac{S_{(k-1)\Delta} - S_{(k-2)\Delta}}{\widetilde{F}_{(k-2)\Delta}}\right)^2 + \left(\frac{S_{k\Delta} + D_{k\Delta} - S_{(k-1)\Delta}}{\widetilde{F}_{(k-1)\Delta}}\right)^2 + \\ + \left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{\widetilde{F}_{k\Delta}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{\widetilde{F}_{T-\Delta}}\right)^2. \qquad (26)$$

Notice that the dividend $D_{k\Delta}$ enters the definition in the term whose denominator is $\widetilde{F}_{(k-1)\Delta}$, but not in the subsequent term. This definition ensures that if the stock price happens to track the forward price at all points in time, then the payoff (26) will be zero in the $\Delta \to 0$ limit, as is the case with variance swaps and simple variance swaps in the absence of dividends.

The starting point of the replicating strategy will be to carry out precisely the trades listed in Tables 2 and 3 with δ set equal to zero (and replacing F_t with $\widetilde{F}_t$ wherever it occurs in the tables). As is easily checked, this replicating strategy generates the payoff in (26) minus V, plus an extra payoff of $(D_{k\Delta}/\widetilde{F}_{(k-1)\Delta})^2 - 2D_{k\Delta}(S_{k\Delta} + D_{k\Delta})/\widetilde{F}^2_{(k-1)\Delta}$. To offset this, two new positions are required: (i) a short position of $e^{-rT}(D_{k\Delta}/\widetilde{F}_{(k-1)\Delta})^2$ (measured in dollars) in bonds, and (ii) a long position of $2D_{k\Delta}e^{-r(T-k\Delta)}/\widetilde{F}^2_{(k-1)\Delta}$ units of the underlying held until time $k\Delta$, then rolled into bonds.

After some algebra (and up to terms of order Δ, as usual) this implies that the simple variance swap strike is given by

$$V = \frac{2e^{rT}}{\widetilde{F}_T^2}\left\{\int_0^{F_T} \mathsf{put}_T(K)\,dK + \int_{F_T}^{\infty} \mathsf{call}_T(K)\,dK\right\},$$

and that the replicating portfolio is equivalent to holding

(i) a static position of $2/\widetilde{F}_T^2$ puts expiring at time T with strike K, for each $K \le F_T$;

(ii) a static position of $2/\widetilde{F}_T^2$ calls expiring at time T with strike K, for each $K \ge F_T$; and

(iii) a dynamic position of $2(F_t - S_t)/(\widetilde{F}_t\widetilde{F}_T)$ units of the underlying asset at time t.