"Simple Variance Swaps"

Ian MARTIN
Stanford University, Graduate School of Business

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 more robust relative of the variance swap that can be priced and hedged even if the underlying asset's price can jump, and constructs SVIX, an index based on simple variance swaps that measures market volatility. SVIX is consistently lower than VIX in the data, which rules out the possibility that the market return and stochastic discount factor are conditionally lognormal. The SVIX series implies a lower bound on the one-month equity premium that peaked at 55%, in annualized terms, at the height of the credit crisis.
Simple Variance Swaps

Ian Martin*

June, 2012

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 more robust relative of the variance swap that can be priced and hedged even if the underlying asset’s price can jump, and constructs SVIX, an index based on simple variance swaps that measures market volatility. SVIX is consistently lower than VIX in the data, which rules out the possibility that the market return and stochastic discount factor are conditionally lognormal. The SVIX series implies a lower bound on the one-month equity premium that peaked at 55%, in annualized terms, at the height of the credit crisis.

Keywords: variance swap, VIX, robust hedging, entropy, equity premium, jumps, implied correlation.

*Stanford GSB and NBER; http://www.stanford.edu/~iwrm/. First draft: November 15, 2010. I thank Torben Andersen, Andy Atkeson, John Campbell, Peter Carr, Mike Chernov, John Cochrane, Darrell Duffie, Bob Hall, Stefan Hunt, Chris Jones, Stefan Nagel, Monika Piazzesi, Steve Ross, Myron Scholes, Costis Skiadas, Andreas Stathopoulos, Viktor Todorov, and seminar participants at USC, at the SITE 2011 conference, and at the Kellogg Finance 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 . . . . 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 Section 1, I define and analyze the simple variance swap of the title, which is robust to the issues mentioned in the quotation above. Simple variance swaps are simple in two senses. First, they 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. Second, they measure the risk-neutral variance of simple returns.

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 VIX index is based on the strike of a variance swap; I introduce an analogous index, SVIX, that is based on the strike of a simple variance swap. (Although the motivation for the definitions of VIX and SVIX comes from the theory of variance swap and simple variance swap pricing, Sections 2 and 3 can be read independently of Section 1.) In the time series from January 1996 to January 2012, VIX
is always higher than SVIX, and the gap between the two—an index of non-lognormality—spikes at times of market stress. In any conditionally lognormal model VIX would be lower than SVIX, so this is model-free evidence that we do not live in a lognormal world. It is considerably stronger than the standard observation that empirical log return distributions are non-Normal, since that leaves open the possibility that log returns are Normal conditional on some information observed by market participants. My approach is robust to critiques based on conditioning information, to the extent that such information is embedded in option prices.

I present various results about the properties of VIX and SVIX. In particular I show, under weak assumptions, that SVIX provides a lower bound for the forward-looking equity premium. I construct the SVIX index and the implied equity premium lower bounds using option price data from OptionMetrics. The results are striking. Although the average lower bound in the data is on the order of 5%, it is extraordinarily variable. The lower bound on the one-month equity premium peaked at 55.0% (annualized) at the height of the 2008–9 crisis, and the lower bound on the one-year equity premium peaked at 21.5%.\footnote{Bollerslev, Tauchen and Zhou (2009), Drechsler and Yaron (2011), and Bekaert and Engstrom (2011) present related empirical results in the context of specific equilibrium models. My approach provides a theoretical reason to expect that such results should hold more generally. It differs from these authors’ approaches in that it does not require estimation of the forward-looking real-world variance of returns.}

Lastly, I show how to use simple variance swaps to calculate market-implied correlations. Simple variance swaps are perfectly suited for this purpose, because the variance of the simple return on the market decomposes nicely as a weighted sum of variances and covariances of simple returns on the individual assets that make up the index.

Related literature. Carr and Corso (2001) and Bondarenko (2007) have proposed different ways around the problem of hedging variance swaps in the presence of jumps. In both cases, the approach only works if the underlying asset is a futures contract. For various reasons, it is desirable to extend beyond this case. It is not currently practical to base the payoffs on single-name simple variance swaps on futures, because the futures market on individual stocks is...
illiquid in the US. The economic interpretation is also complicated because futures price processes are not true price processes but cumulative dividend processes; thus neither definition provides a measurement of the risk-neutral variance of returns, which is the quantity of interest in Section 3. In that section it is also helpful to think of the asset underlying the simple variance swap as being the portfolio held by the representative agent. This asset cannot be a futures contract, because futures are in zero net supply. Finally, letting the underlying asset be an index permits a direct comparison with VIX, which is based on index options as opposed to index futures options.


1 Variance swaps and simple variance swaps

Today is time 0; unsubscripted expectation operators denote expectations conditional on time-0 information. Consider an underlying asset whose price at time $t$ is $S_t$, and which continuously pays dividends at rate $\delta S_t$ per unit time. In subsequent sections, I will follow the related literature in assuming that the asset does not pay dividends, $\delta = 0$; 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” $\delta$ 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.) I require some other assumptions that, again, are standard in the related literature: (i) that the interest rate is constant, at $r$ (an assumption that will be relaxed in subsequent sections); (ii)

\[\text{See Duffie (2001, pp. 171–172).}\]
that there is a frictionless market in European puts and calls on the underlying asset; (iii) that there are no transactions costs when trading the underlying or options on it; (iv) that there is no counterparty default; and (v) that the asset can be traded at the appropriate frequency to permit rebalancing of the replicating portfolio.

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

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

(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 $\Delta$ 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 $\text{call}_t(K)$ for the time-0 price of a European call option on the underlying asset, expiring at time $t$, with strike $K$, and $\text{put}_t(K)$ 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

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

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

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

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.,

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

We have

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

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

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\(^3\)

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

\(^3\)Neuberger (1990), the working paper version of Neuberger (1994), mentions the relevance of this “Squared contract” in a similar context.
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

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

The right-hand side is the time-\( t \) payoff on an equal-weighted portfolio of European call options of all strikes, so the absence of arbitrage implies that

\[
\Pi(t) = 2 \int_0^\infty \text{call}_t(K) \, dK.
\]

Equation (3) follows by put-call parity, which is the relationship \( \text{call}_t(K) = \text{put}_t(K) + e^{-rt}(F_t - K) \). 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.

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 \( \Delta, 2\Delta, \ldots, T \). But, correspondingly, the hedging portfolio requires holding portfolios of options of each of these maturities. Although this is not a serious issue if \( \Delta \) is large relative to \( T \), it raises the concern that hedging a simple variance swap may be extremely costly in practice if \( \Delta \) 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.

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

\[
V(0,T) \equiv \lim_{\Delta \to 0} V(\Delta, T) = \frac{e^{rT}P(T)}{F_T^2},
\]

where

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

\(^4\)Darrell Duffie suggested this approach. A previous draft derived (5) via Breeden-Litzenberger (1978) logic and integration by parts: \( \Pi(t) = \int_0^{\infty} K^2 \text{call}_t''(K) \, dK = 2 \int_0^{\infty} \text{call}_t(K) \, dK \). The latter approach is less neat, but has the advantage of being mechanical: it does not rely on a "trick".
Proof. Observe that (2) can be rewritten
\[ V(\Delta, T) = \sum_{i=1}^{T/\Delta} \left\{ \frac{e^{ri\Delta}}{F^2_{(i-1)\Delta}} \left[ P(i\Delta) - (2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta}P((i-1)\Delta) \right] \right\} + \frac{T}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2. \]  

(7)

For \( 0 < j < T/\Delta \), the coefficient on \( P(j\Delta) \) in equation (7) is
\[ \frac{e^{rj\Delta}}{F^2_{(j-1)\Delta}} - \frac{e^{r(j+1)\Delta}}{F^2_{j\Delta}}(2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta} = \frac{e^{rj\Delta}}{F^2_{j\Delta}} \left( e^{(r-\delta)\Delta} - 1 \right)^2. \]

(At this point it may be helpful to note that the proof, and the definition (1), were originally found by viewing the normalizing constants \( F_{j\Delta} \), for \( j = 0, \ldots, T/\Delta \), as arbitrary, and choosing them so that the above equation would hold.) We can therefore rewrite (7) as
\[ V(\Delta, T) = \frac{e^{rT}}{F^2_{T-\Delta}} P(T) + \sum_{j=1}^{T/\Delta-1} \frac{2e^{rj\Delta}}{F^2_{j\Delta}} \left( e^{(r-\delta)\Delta} - 1 \right)^2 P(j\Delta) + \frac{T}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2. \]

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 \( \Delta^2 \); 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 \( e^{rT}P(T)/F^2_{T} \), 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.\(^5\)

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

\(^5\)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\%.
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 $\Delta$ 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 $\Delta$, as in equation (5); this portfolio has simple return $S_\Delta^2/\Pi(\Delta)$ from time 0 to time $\Delta$. The fifth, sixth, and seventh rows indicate how the proceeds of this option portfolio are used after time $\Delta$. Part of the proceeds are immediately invested in the bond until time $T$; another part is invested from $\Delta$ to $2\Delta$ in the underlying asset, and subsequently from $2\Delta$ 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 $\Delta$ to time $T-\Delta$ is zero. The last column of Table 2 adds up to the desired payoff,

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

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 $2e^{-r(T-(j+1)\Delta)}S_j^2e^{-\delta\Delta}/F_j^2$ in dollar terms at time $j\Delta$, i.e. a short position of $2e^{-r(T-(j+1)\Delta)}S_j^2e^{-\delta\Delta}/F_j^2$ 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 $t$. 

9
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 $F_t = S_0e^{(r-\delta)t}$, 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

1. a static position in $2/F_T^2$ puts expiring at time $T$ with strike $K$, for each $K \leq F_T$;
2. a static position in $2/F_T^2$ calls expiring at time $T$ with strike $K$, for each $K \geq F_T$; and
3. 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 $\Delta$ 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 $\delta S_t$, 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 follow the literature in assuming, for simplicity, that the underlying asset does not pay dividends, $\delta = 0$.

In contrast, a (standard) variance swap pays

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

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

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

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, $\tilde{V}$ 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*

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

which has the interpretation

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

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 $K \leq F_T$;

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

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

\footnote{Jarrow et al. (2010) provide technical conditions under which $\tilde{V}$ converges to $\mathbb{E}^*[(\log S)_T]$, where $(\log S)_t$ is the quadratic variation process of $\log S_t$.}
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

\[
\tilde{V} = \mathbb{E}^* \left[ \int_0^T \sigma_t^2 dt \right] = 2 \mathbb{E}^* \left[ \int_0^T \frac{1}{S_t} dS_t - \int_0^T d \log S_t \right] = 2rT - 2 \mathbb{E}^* \log \frac{S_T}{S_0}.
\]

(11)

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 \( R_T = \frac{S_T}{S_0} \). 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:

\[
P_{\log} \equiv e^{-rT} \mathbb{E}^* \log R_T = rT e^{-rT} - \int_0^{F_T} \frac{1}{K^2} \text{put}_T(K) dK - \int_{F_T}^\infty \frac{1}{K^2} \text{call}_T(K) dK.
\]

(12)

Substituting (12) back into (11), we have the result.

There is a telling contrast between the hedging portfolios of variance swaps and simple variance swaps. 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, reflecting the fact that variance swaps are more sensitive than simple variance swaps to downward jumps in the price of the underlying asset.

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). In contrast with the hedge portfolio for a standard variance swap, which holds 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.

![Sample path](image)

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).

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.\(^7\) In Panel 1b, only one line is visible: the simple variance swap is essentially perfectly hedged. In contrast, Panel 1e shows that the hedge portfolio

\(^7\)For times $t$ prior to expiry, I compute the required payout and hedge portfolio perfor-
(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.⁸

2 VIX and SVIX

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

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

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, given in (6):

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

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.

Although the original motivation behind the definitions of VIX and SVIX comes from the variance swap and simple variance swap contracts, it will be
convenient from now on simply to take them as given. The benefit of doing so is that in the remainder of the paper we will be able to drop the assumption that the interest rate is constant. From now on, I write \( r \) for the \( T \)-period zero-coupon yield that prevails at time 0, and there is no presumption that interest rates are constant after time 0.

Under the Itô process assumption, \( \text{VIX}^2 \) corresponds to the (annualized) strike on a variance swap, and has the interpretation (10). If there are jumps, though, the correctly priced strike \( \tilde{V} \) 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 \( \tilde{V} \) nor \( \text{VIX}^2 \) has the interpretation (10). The next result shows what \( \text{VIX} \) does measure, and contrasts this with the much simpler interpretation of \( \text{SVIX} \).

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

\[
\text{VIX}^2 = \frac{2}{T} \text{L}^*(R_T),
\]

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

\[
\text{VIX}^2 = \frac{1}{T} \text{var}^* \log R_T \approx \frac{1}{T} \text{var}^* R_T,
\]

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

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

*where \( \kappa_n^* \equiv \frac{1}{T} \tilde{\kappa}_n^* \) and \( \tilde{\kappa}_n^* \) is the \( n \)th cumulant of \( \log R_T \).*

*In contrast, \( \text{SVIX} \) measures the risk-neutral variance of the simple return:*

\[
\text{SVIX}^2 = \frac{1}{T} \text{var}^* R_T.
\]

\[9\text{So, for example, } \tilde{\kappa}_1^* = \mathbb{E}^* \log R_T; \tilde{\kappa}_2^* = \text{var}^* \log R_T; \tilde{\kappa}_3^* \text{ is the skewness of } \log R_T \text{ multiplied by } (\tilde{\kappa}_2^*)^{3/2}; \tilde{\kappa}_4^* \text{ is the excess kurtosis, multiplied by } (\tilde{\kappa}_2^*)^2; \text{ and so on.} \]
Proof. Equation (14) follows from the definition of $\text{VIX}^2$ and (12), together with the fact that $\mathbb{E}^* R_T = e^{rT}$.

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

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

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

where $\bar{\kappa}_n^*$ is the $n$th risk-neutral cumulant of $\log R_T$. The definition of entropy implies that $L^*(R_T) = \kappa^*(1) - \kappa^*(0)$, from which (16) follows after annualizing the cumulants: $\kappa_n^* \equiv \frac{1}{T} \bar{\kappa}_n^*$. 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

$$\text{var}^* R_T = \mathbb{E}^* \left[ \left( \frac{S_T}{S_0} \right)^2 \right] - \left[ \mathbb{E}^* \left( \frac{S_T}{S_0} \right) \right]^2 = e^{rT} \Pi(T) \frac{S_2}{S_0} - e^{2rT}.$$

From (3), this implies

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

from which (17) follows. \qed

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 $L(\cdot)$ 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
option prices

![Diagram](image)

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 (2011).\footnote{Entropy is an overloaded term. I use it here because of the link to Theil (1967). Backus, Chernov and Martin (2011) refer to $L(M)$ as the entropy of a stochastic discount factor $M$ because in a complete market $L(M)$ equals the relative entropy, in the information-theoretic sense, of risk-neutral probabilities with respect to real-world probabilities.}

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 returns $R_T^{(i)}$, $i = 1, 2, \ldots$, and arrives at the overall portfolio return $R_T$. In other words, he chooses portfolio weights $\{w_i\}$ to solve the maximization problem

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

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 $\gamma$, then VIX can be expressed in terms of the cumulants of $\log R_T$ under the real-world probabilities, i.e. $\kappa_2 = \frac{1}{T} \text{var} \log R_T$ and so on:

$$VIX^2 = \sum_{n=2}^{\infty} \alpha_n \kappa_n, \quad (19)$$

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

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

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 $\kappa^*(\theta) = \kappa(\theta - \gamma) - \kappa(-\gamma)$. Equation (20) becomes

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

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

To interpret this result, note that if $\gamma \geq 1$, then $\alpha_n$ 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 riskless return from time 0 to time \( T \). The result requires that any one of the following assumptions holds:

**A1** There is a one-period marginal 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 marginal investor with separable utility whose value function \( J \), which depends on current wealth \( W_0 \), can be defined recursively as

\[
J[W_0] = \max_{C_0, \{w_i\}} w(C_0) + \beta \mathbb{E} \left[(W_0 - C_0) \sum_i w_i R^{(i)}_T \right] \quad \text{s.t.} \quad \sum_i w_i = 1,
\]

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

**A3** There is an Epstein-Zin (1989) marginal investor with constant consumption-wealth ratio and risk aversion \( \gamma \geq 1 \).

**A4** The SDF and market return are conditionally jointly lognormal, and the market’s conditional Sharpe ratio exceeds its conditional volatility.

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 Wachter’s (2011) time-varying disaster risk model.\(^{11}\) Finally, Wachter’s model is in continuous time; Assumption A3 applies to a discretization in which the investor consumes at times 0, \( T \), \( 2T \), \ldots, and so on.
Assumption A4 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.

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

\[
\frac{1}{T} \mathbb{E} [R_T - R_{f,T}] \geq \frac{SVIX^2}{R_{f,T}}. \tag{21}
\]

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

**Proof.** If there is a representative investor with log utility then \(1/R_T\) is an SDF, so that \(1/R_{f,T} \mathbb{E}^* X = \mathbb{E}[X/R_T]\) 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, or A3, we will show that \(M_T R_T\) is decreasing in \(R_T\), from which the result follows. Assumption A4 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 A3 holds, then the SDF is equal (up to constants of proportionality that are known at time 0) to \( C_T^{\theta / \psi} R_T^{\vartheta - 1} \), where \( \vartheta \equiv (1 - \gamma )/(1 - 1/\psi) \), with \( \gamma \) denoting the investor’s risk aversion and \( \psi \) denoting the elasticity of intertemporal substitution; and the representative investor’s wealth is proportional to the return on the market. Together, these imply that

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

which gives the result.

If A4 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 \( \lambda \geq \sigma_R \).

That is, SVIX provides a bound on 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:

\[
\frac{\text{var}^* R_T}{R_{f,T}} \leq \mathbb{E} R_T - R_{f,T} \leq R_{f,T} \cdot \sigma (M) \cdot \sigma (R_T).
\]

The left inequality is (22); the right inequality is the Hansen-Jagannathan bound, with \( M \) denoting the SDF to time \( T \) and \( \sigma (\cdot) \) denoting conditional standard deviations computed with respect to the real-world probabilities. To exploit the Hansen-Jagannathan bound, time-series averages are typically used to compute backward-looking Sharpe ratios as a proxy for the true, but unobservable, measure of interest, the forward-looking Sharpe ratio. The important feature of (22) is that \( \text{var}^* R_T \), which implies a lower bound on the equity premium, 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

\[ \text{instantaneous risk premium} = \gamma \sigma^2 \]  

(23)

where \( \gamma \) is a measure of aggregate risk aversion, and \( \sigma^2 \) 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 \( \gamma \).

There are some important differences between the two approaches. In particular, Merton assumes that the market’s price follows a diffusion, thereby ruling out the effects of skewness and of higher moments by construction.\(^{12}\) 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 \( \frac{R_{t,T}}{T} \mathbb{E}[R_T - R_{f,T}] = \gamma \cdot \text{SVIX}^2 \). 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 \( \frac{R_{t,T}}{T} \mathbb{E}[R_T - R_{f,T}] \geq \gamma \cdot \text{SVIX}^2 \). 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

\(^{12}\)Amongst others, Rubinstein (1973), Kraus and Litzenberger (1976), and Harvey and Siddique (2000) emphasize the importance of skewness in portfolio choice.
could be calculated “by ‘inverting’ the Black-Scholes option pricing formula”. However, Black-Scholes implied volatility would only provide the correct measure of $\sigma$ 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

I construct the time series of SVIX using option price data supplied by OptionMetrics, following the methodology used to construct VIX. 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.

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

Figures 3a and 3b plot SVIX from January 4, 1996, to January 31, 2012 with $T = 1$ month. For clarity, all figures show 10-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 $R_T$

\[^{13}\text{In principle—unfortunately index options were not traded when he wrote the paper.}\]

\[^{14}\text{See The CBOE Volatility Index – VIX, available on the CBOE website.}\]
were lognormally distributed under the risk-neutral measure, as the next result will show; loosely 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 \( \frac{1}{T} (2L^* (R_T) - \text{var}^* R_T) \).

Unsurprisingly, therefore, the VIX-SVIX spread jumped in the recent crisis and at other times of market stress.

**Result 8.** If the SDF \( M_T \) and return \( R_T \) are conditionally jointly lognormal, then \( \text{SVIX}^2 = \frac{1}{T} e^{2rT} (e^{\sigma^2_T} - 1) \) and \( \text{VIX}^2 = \sigma_R^2 \), where \( \sigma_R^2 = \frac{1}{T} \text{var} \text{log} R_T \).

**Proof.** Let \( R_T = e^{\mu_T + \sigma_T \sqrt{T} Z_R - \sigma_R T^2 / 2} \) and \( M_T = e^{-rT + \sigma_M \sqrt{T} Z_M - \sigma_M^2 T^2 / 2} \), where \( Z_R \) and \( Z_M \) are Normal random variables with mean zero, variance one, and correlation \( \rho \). Since \( \mathbb{E} M_T R_T = 1 \), we must have \( \mu_T - r = -\rho \sigma_M \sigma_R \).

From (17), \( \text{SVIX}^2 = \frac{1}{T} \text{var}^* R_T = \frac{1}{T} \left[ (\mathbb{E}^* (R_T^2) - (\mathbb{E}^* R_T)^2) \right] = \frac{1}{T} (e^{2rT} \mathbb{E} M_T R_T^2 - e^{2rT}) \).

Now, using the fact that \( \mu_T - r = -\rho \sigma_M \sigma_R \), we have

\[
\mathbb{E} M_T R_T^2 = \mathbb{E} e^{-rT + \sigma_M \sqrt{T} Z_M - \frac{1}{2} \sigma_R^2 T} e^{2\mu_T + 2\sigma_R \sqrt{T} Z_R - \sigma_R^2 T} = e^{(2\mu_T - r)T + \frac{1}{2} \sigma_R^2 T} e^{2\sigma_M \sigma_R T} = e^{rT + \frac{1}{2} \sigma_R^2 T}.
\]

Thus, \( \text{SVIX}^2 = \frac{1}{T} e^{2rT} (e^{\sigma^2_R T} - 1) \) as required.

The calculation for VIX is slightly more complicated. Using (14), \( \text{VIX}^2 = \frac{2}{T} R_T = \frac{2}{T} \left[ \log \mathbb{E}^* R_T - \mathbb{E}^* \text{log} R_T \right] = \frac{2}{T} \left[ rT - e^{rT} \mathbb{E} M_T \text{log} R_T \right] \). Now,

\[
\mathbb{E} [M_T \text{log} R_T] = \mathbb{E} \left[ (\mu_T \text{log} (e^{2\mu_T} - \frac{1}{2} \sigma_R^2 T) e^{-rT + \sigma_M \sqrt{T} Z_M - \sigma_R^2 T}) \right]
= (\mu_T - \frac{1}{2} \sigma_R^2 T) e^{-rT + \sigma_M \sqrt{T} Z_M - \sigma_R^2 T} \mathbb{E} \left[ Z_R e^{\sigma_M \sqrt{T} Z_M} \right].
\]

We can write \( Z_R = \rho Z_M + \sqrt{1 - \rho^2} Z \), where \( Z \) is uncorrelated with \( Z_M \) and hence, since they are both Normal, independent of \( Z_M \). The expectation in the above expression then becomes

\[
\mathbb{E} \left[ Z_R e^{\sigma_M \sqrt{T} Z_M} \right] = \mathbb{E} \left[ (\rho Z_M + \sqrt{1 - \rho^2} Z) e^{\sigma_M \sqrt{T} Z_M} \right] = \rho \mathbb{E} \left[ Z_M e^{\sigma_M \sqrt{T} Z_M} \right].
\]

By Stein’s lemma,

\[
\mathbb{E} \left[ Z_M e^{\sigma_M \sqrt{T} Z_M} \right] = \mathbb{E} \left[ \sigma_M \sqrt{T} e^{\sigma_M \sqrt{T} Z_M} \right] = \sigma_M \sqrt{T} e^{\sigma^2_M T / 2}.
\]
These results, together with the fact that $\mu_R - r = -\rho \sigma_M \sigma_R$, imply that
\[\mathbb{E} M_T \log R_T = (\mu_R - \frac{1}{2} \sigma_R^2 + \rho \sigma_M \sigma_R) T e^{-rT} = (r - \frac{1}{2} \sigma_R^2) T e^{-rT}.\]
It follows that $VIX^2 = \frac{2}{T} \left[ rT - (r - \frac{1}{2} \sigma_R^2) T \right] = \sigma_R^2$, as required.

Since $e^{\sigma_R^2 T} - 1 > \sigma_R^2 T$, $VIX$ is lower than $SVIX$ in any conditionally log-normal model. The opposite is true in the data, as Figure 3b shows. This is direct, model-free evidence that at the 1-month horizon, returns and the SDF are not conditionally lognormal. Moreover, $VIX$ is higher than $SVIX$ on every single day in my sample. It is not that nonlognormality only matters at times of crisis: it is a completely pervasive feature of the data.\(^{15}\)

This is consistent with the findings of Bollerslev and Todorov (2011a, 2011b).\(^{16}\) Bollerslev and Todorov require identifying assumptions\(^{17}\) to draw their conclusions from high-frequency data, however, whereas my conclusion is based on a direct comparison of the price of one portfolio of options to another. Since none of my results requires estimation of quantities under the true ($\mathbb{P}$) measure, no identifying assumptions are needed.

It is worth emphasizing that this evidence is much stronger than the now familiar observation that histograms of log returns are not (approximately) Normal, since that leaves open the possibility that log returns are conditionally Normal. Put differently, it is possible in principle that conditionally Normal log returns generate highly non-Normal unconditional log return distributions. Figure 3b excludes that possibility. No model in which the market return and SDF are conditionally jointly lognormal is consistent with the data.\(^{15}\)

\(^{15}\)Figure 5 in the appendix shows the corresponding plots for 3-month, 6-month, and 1-year horizons. In each case, $VIX$ has almost invariably been higher than $SVIX$.

\(^{16}\)Several authors find similar results in parametric frameworks: for example, Bates (2000), Ait-Sahalia et al. (2001), Andersen et al. (2002), Pan (2002), Carr and Wu (2003), Eraker (2004), Broadie et al. (2009), and Backus et al. (2011).

\(^{17}\)Specifically, they make assumptions that link the time-variation of arrival rates of severe disasters to the arrival rates of small disasters, and that restrict the distribution of extreme jumps. See Assumptions (2.6), (3.6), (A.2), and the assumptions in the statement of Proposition 1, in Bollerslev and Todorov (2011a).
3.1 A lower bound on the equity premium

Figure 4a plots the 10-day moving average of $\frac{1}{R_{f,T}}$SVIX$^2$ measured in percentage points, for the one-month horizon. 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. During the years 2004–2006, the average lower bound was only 1.86%; in contrast, during the recent crisis, the lower bound peaked at 38.7% on a 10-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, and 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 also around May 2010, coinciding with the beginning of the European sovereign debt crisis. Other 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 a 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, the lower bound has the appealing property that, by construction, it can never be less than zero.

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
Figure 4: The lower bound on the annualized equity premium at different horizons (in %). The figures show 10-day moving averages.
Table 1: Mean, standard deviation, and quantiles of equity premium bounds at various horizons (annualized and measured in %).

<table>
<thead>
<tr>
<th>horizon</th>
<th>mean</th>
<th>s.d.</th>
<th>min</th>
<th>1%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>99%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mo</td>
<td>5.00</td>
<td>4.60</td>
<td>0.83</td>
<td>1.03</td>
<td>1.54</td>
<td>2.44</td>
<td>3.91</td>
<td>5.74</td>
<td>8.98</td>
<td>25.7</td>
<td>55.0</td>
</tr>
<tr>
<td>2 mo</td>
<td>5.00</td>
<td>3.99</td>
<td>1.01</td>
<td>1.20</td>
<td>1.65</td>
<td>2.61</td>
<td>4.11</td>
<td>5.91</td>
<td>8.54</td>
<td>23.5</td>
<td>46.1</td>
</tr>
<tr>
<td>3 mo</td>
<td>4.96</td>
<td>3.60</td>
<td>1.07</td>
<td>1.29</td>
<td>1.75</td>
<td>2.69</td>
<td>4.24</td>
<td>5.95</td>
<td>8.17</td>
<td>21.4</td>
<td>39.1</td>
</tr>
<tr>
<td>6 mo</td>
<td>4.89</td>
<td>2.97</td>
<td>1.30</td>
<td>1.53</td>
<td>1.95</td>
<td>2.88</td>
<td>4.39</td>
<td>6.00</td>
<td>7.69</td>
<td>16.9</td>
<td>29.0</td>
</tr>
<tr>
<td>1 yr</td>
<td>4.63</td>
<td>2.43</td>
<td>1.22</td>
<td>1.64</td>
<td>2.07</td>
<td>2.79</td>
<td>4.35</td>
<td>5.71</td>
<td>7.19</td>
<td>13.9</td>
<td>21.5</td>
</tr>
</tbody>
</table>

equity premium bound is SVIX squared, which in turn is more skewed and has a higher kurtosis than VIX.

### 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 $\sigma^2_M$ for the risk-neutral variance of the simple return on the market from time 0 to time $T$; write $\sigma^2_i$ for the risk-neutral variance of the simple return on stock $i$ from time 0 to time $T$; write $\rho_{ij}$ for the correlation between stocks $i$ and $j$; and write $w_i$ for the market weights:

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

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

$$\hat{\rho} = \frac{\sigma^2_M - \sum_{i \neq j} w_i^2 \sigma_i^2}{\sum_{i \neq j} w_i w_j \sigma_i \sigma_j}. $$
The implied correlation \( \hat{\rho} \) can then be calculated directly from the strike on a simple variance swap on the market (which reveals \( \sigma_M^2 \), via equations (6) and (17)) and from the strikes on simple variance swaps on the index constituents (which reveal \( \{\sigma_i^2\}_{i=1,...,N} \), 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 lognormal 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, as 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.

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 simple interpretation: whether or not there are jumps, they reveal the risk-neutral variance of the simple return on the underlying asset.

The definition of a simple variance swap can be used to define an index, SVIX, that is analogous to VIX. In the time series from January 1996 to January 2012, SVIX is always lower than VIX, and the gap between the two—an index of non-lognormality—spikes up at times of market stress. This is direct evidence that we do not live in a lognormal world.

Simple variance swaps have a natural application to measuring implied correlation. They can also be used to derive a lower bound on the forward-looking expected equity premium. 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 5.00% in annualized terms for returns over a one-month horizon and 4.63% for a one-year horizon—is reassuringly close to the long-run average realized equity premium. But it is also extraordinarily volatile: at the height of the recent crisis, in November 2008, the lower bound on the one-year equity premium rose to 21.5%, and the lower bound on the annualized one-month equity premium climbed to 55.0%.

5 Bibliography


### Appendix

**Equation (6) is an excellent approximation even if $\Delta > 0$**

In this section I show that (6), which holds exactly in the $\Delta \to 0$ limit, is an excellent approximation to the truth away from the limit point for all reasonable values of $\Delta$. I do so 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 9.** For $\Delta > 0$, $V(\Delta, T)$ is very well approximated by $\frac{e^{rT}P(T)}{F_{T-\Delta}}$:

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

(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 \( \Delta \) with \( t_1 \leq t_2 \), we have
\[
\frac{e^{rt_1} P_{t_1}}{F_{t_1}^2} = \lim_{\Delta \to 0} \mathbb{E}^* \left( \sum_{i=1}^{t_1/\Delta} \left[ \frac{S_{i\Delta} - S_{(i-1)\Delta}}{F_{(i-1)\Delta}} \right]^2 \right) \leq \lim_{\Delta \to 0} \mathbb{E}^* \left( \sum_{i=1}^{t_2/\Delta} \left[ \frac{S_{i\Delta} - S_{(i-1)\Delta}}{F_{(i-1)\Delta}} \right]^2 \right) = \frac{e^{rt_2} P(t_2)}{F_{t_2}^2},
\]
so that \( e^{rt} P(t)/F_t^2 \) is increasing in \( t \). It follows that
\[
\left| V(\Delta, T) - \frac{e^{rT} P(T)}{F_{T-\Delta}^2} \right| = \sum_{j=1}^{T/\Delta - 1} \left( e^{(r-\delta)\Delta} - 1 \right)^2 \frac{e^{rj\Delta} P(j\Delta)}{F_{j\Delta}^2} + \frac{T}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2
\leq \left( \frac{T}{\Delta} - 1 \right) \left( e^{(r-\delta)\Delta} - 1 \right)^2 \frac{e^{rT} P(T)}{F_T^2} + \frac{T}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2
\leq \left( \frac{T}{\Delta} - 1 \right) \left( e^{(r-\delta)\Delta} - 1 \right)^2 \frac{e^{rT} P(T)}{F_{T-\Delta}^2} + \frac{T}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2
\leq T \left( e^{(r-\delta)\Delta} - 1 \right)^2 \cdot \left( 1 + \frac{e^{rT} P(T)}{F_{T-\Delta}^2} \right),
\]
as required. \( \square \)

This result has bite because the term \( T \left( e^{(r-\delta)\Delta} - 1 \right)^2 \) 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 0.05\( T \). (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
\[
\frac{T}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2 \cdot \left( 1 + \frac{e^{rT} P(T)}{F_{T-\Delta}^2} \right) \approx \frac{20}{\Delta} \left( e^{(r-\delta)\Delta} - 1 \right)^2 (1 + 0.05T).
\]
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\%. 

34
Construction of SVIX

The data are from OptionMetrics, running from January 4, 1996, to January 31, 2012; 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 1,165,585 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

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

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 $T = 30, 60, 90, 180, \text{and } 360$ 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.
<table>
<thead>
<tr>
<th>asset</th>
<th>0</th>
<th>$\Delta$</th>
<th>$2\Delta$</th>
<th>...</th>
<th>$T - \Delta$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>$-e^{-rT}$</td>
<td>$-2e^{-r(T-\Delta)}$</td>
<td>...</td>
<td>$S_0^3 / S_0$</td>
<td>$S_0^3 / S_0$</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>$2e^{-r(T-\Delta)}e^{-\delta \Delta}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>$2e^{-r(T-\Delta)}S_{2\Delta}/S_0$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>$-2S_0S_{2\Delta}/S_0$</td>
<td></td>
</tr>
<tr>
<td>(\Delta)</td>
<td>$-\frac{(e^{(r-\delta)\Delta}-1)^2 \Pi_\Delta}{e^{r(T-\Delta)F^2_\Delta}}$</td>
<td>$\frac{(e^{(r-\delta)\Delta}-1)^2 S_0^3}{e^{r(T-\Delta)F^2_\Delta}}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>$-e^{-r(T-\Delta)} \left( \frac{S_0^3}{S_0^3} + \frac{S_0^3}{F^2_\Delta} \right)$</td>
<td>...</td>
<td>...</td>
<td>$\frac{S_0^3}{F^2_\Delta} + \frac{S_0^3}{F^2_\Delta}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>$\frac{2e^{-r(T-2\Delta)}S_{2\Delta}^2}{F^2_\Delta} e^{-\delta \Delta}$</td>
<td>$-2e^{-r(T-2\Delta)}S_{2\Delta}S_{2\Delta}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>$\frac{2e^{-r(T-2\Delta)}S_{2\Delta}^2}{F^2_\Delta} S_{2\Delta} S_{2\Delta}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>$-2S_{2\Delta}S_{2\Delta}/F^2_\Delta$</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>$T - \Delta$</td>
<td>$-\frac{(e^{(r-\delta)\Delta}-1)^2 \Pi_{T-\Delta}}{e^{r(T-\Delta)F^2_{T-\Delta}}}$</td>
<td>...</td>
<td>...</td>
<td>$\frac{(e^{(r-\delta)\Delta}-1)^2 S_{T-\Delta}^3}{e^{r(T-\Delta)F^2_{T-\Delta}}}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>...</td>
<td>$-e^{-r\Delta} \left( \frac{S_{T-\Delta}^3}{F^2_{T-\Delta}} + \frac{S_{T-\Delta}^3}{F^2_{T-\Delta}} \right)$</td>
<td>$\frac{S_{T-\Delta}^3}{F^2_{T-\Delta}} + \frac{S_{T-\Delta}^3}{F^2_{T-\Delta}}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>...</td>
<td>$\frac{2S_{T-\Delta}^2}{F^2_{T-\Delta}} e^{-\delta \Delta}$</td>
<td>$-2S_{T-\Delta}S_{T-\Delta}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td>$-\frac{\Pi_{T-\Delta}}{F^2_{T-\Delta}}$</td>
<td>...</td>
<td>...</td>
<td>$\frac{S_{T-\Delta}^3}{F^2_{T-\Delta}}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>$V e^{-rT}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>$-V$</td>
<td></td>
</tr>
</tbody>
</table>

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 \(j = 1, 2, \ldots, T/\Delta\), 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 \(\Pi_{j\Delta}\).
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 $S^2_i$ 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 $\tilde{F}_t \equiv S_0 e^{rt}$, 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

$$F_t = \begin{cases} 
S_0 e^{rt} & \text{if } t < k\Delta \\
nS_0 e^{rt} - D_{k\Delta} e^{r(t-k\Delta)} & \text{if } t \geq k\Delta 
\end{cases}$$

so $F_t$ and $\tilde{F}_t$ coincide for times $t$ before the payment of the dividend, but differ

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 $j\Delta$.
thereafter. It turns out that $\tilde{F}_t$ 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

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

Notice that the dividend $D_{k\Delta}$ enters the definition in the term whose denominator is $\tilde{F}_{(k-1)\Delta}$, but not in the subsequent term. This convention 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 $\delta$ set equal to zero (and replacing $F_t$ with $\tilde{F}_t$ 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 $(D_{k\Delta}/\tilde{F}_{(k-1)\Delta})^2 - 2D_{k\Delta}(S_{k\Delta} + D_{k\Delta})/\tilde{F}_{(k-1)\Delta}^2$. To offset this, two new positions are required: (i) a short position of $e^{-rT}(D_{k\Delta}/\tilde{F}_{(k-1)\Delta})^2$ (measured in dollars) in bonds, and (ii) a long position of $2D_{k\Delta}e^{-r(T-k\Delta)}/\tilde{F}_{(k-1)\Delta}^2$ units of the underlying held until time $k\Delta$, then rolled into bonds.

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

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

and that the replicating portfolio is equivalent to holding

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

(iii) a dynamic position of $2(F_t - S_t)/(\tilde{F}_t \tilde{F}_T)$ units of the underlying asset at time $t$. 
Figure 5: VIX-SVIX spread over different horizons.