Simple Variance Swaps

Ian Martin*

January, 2013

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 time series, which rules out the possibility that the market return and stochastic discount factor are conditionally lognormal. The SVIX index points to an equity premium that—in contrast to the prevailing view in the literature—is extraordinarily volatile and that spiked dramatically at the height of the recent crisis.

Keywords: variance swap, VIX, equity premium, jumps, entropy.

*Stanford GSB and NBER; http://www.stanford.edu/∼iwrm/. First draft: November 15, 2010. I thank Torben Andersen, Andy Atkeson, Jack Busta, John Campbell, Peter Carr, Mike Chernov, John Cochrane, George Constantinides, Bernard Dumas, Darrell Duffie, Bob Hall, Stefan Hunt, Chris Jones, Stefan Nagel, Anthony Neuberger, Monika Piazzesi, Steve Ross, Myron Scholes, Costis Skiadas, Andreas Stathopoulos, Viktor Todorov; seminar participants at USC, the SITE 2011 conference, the Kellogg Finance Conference, the NBER Summer Institute, INSEAD, LSE, and EPFL/University of Lausanne; and the Editor, Associate Editor, and two anonymous referees 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. Unfortunately, the variance swap market experienced turmoil as the stock market dropped sharply during the credit crisis of 2008–9. The single-name variance swap market was particularly severely affected: it collapsed, and has not recovered.

To explain why, I review the standard theory of variance swap pricing and hedging in Section 1. The fundamental problem is that there is no known way to replicate the payoff of a variance swap if the underlying asset's price can jump. This problem applies with particular force to individual stocks, which are more susceptible to jumps than indices are. The presence of jumps also invalidates the conventional interpretation of the VIX index, so I provide a more general interpretation.

In Section 2, I define and analyze the simple variance swap contract. Simple variance swaps are simple in two senses. First, they are simple to price and hedge: in particular, they can be hedged in the presence of jumps. Second, they measure the risk-neutral variance of simple returns. Simple variance swaps are also robust to several potential concerns regarding practical implementation. Perhaps most important, I show that my hedging and pricing results also hold, to very high accuracy, if monitoring and hedging occurs at discrete points in time, rather than continuously; this is, of course, the case that applies in practice. There is no corresponding result for variance swaps.

Just as VIX is based on the strike of a variance swap, one can consider an index, SVIX, that is based on the strike of a simple variance swap. In Section 3, I construct the time series of SVIX from January 1996 to January 2012 using S&P 500 index option price data from *OptionMetrics*. VIX is higher than SVIX throughout the sample, 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 conditionally lognormal world.

Section 4 connects the SVIX index to the equity premium. I exploit an identity that relates the equity premium to risk-neutral variance and show, under a weak assumption (the *negative correlation condition*), that SVIX provides a lower bound for the forward-looking equity premium. This lower bound has striking properties. Although it averages about 5% over the sample period, it varies dramatically, and at fairly high frequency. It implies that at the height of the 2008–9 crisis, the one-month equity premium was at least 55.0% (annualized), and that the one-year equity premium was at least 21.5%. Again, I address various issues regarding practical implementation, and show that the lower bound is conservative: it would be even higher if option prices were perfectly observable at all strikes.

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). Bakshi, Kapadia and Madan (2003) and Neuberger (2011) show how to use option prices to measure the higher risk-neutral moments of log returns. Lee (2010a, 2010b) provides a brief summary of volatility derivative pricing in the absence of jumps. Carr and Lee (2009) provide an excellent survey of the state of the art in the area.

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 requires the underlying asset to be a futures contract.

Bollerslev, Tauchen and Zhou (2009), Drechsler and Yaron (2011), and Bekaert and Engstrom (2011) relate variance risk premia to equity premia in the context of specific equilibrium models. This paper provides a reason to expect that such results should hold more generally.

Notation. The current date is time 0, and the terminal time horizon is T. Throughout the paper, expectations, variances, and so on, are conditional on current information. Asterisks indicate expectations and variances calculated with respect to the risk-neutral measure. There is an underlying asset whose spot price at time t is S_t and whose forward price to time t—which is known at time 0—is $F_{0,t}$. The time-0 price of a European call option on the underlying asset, expiring at time t, with strike K, is $\text{call}_{0,t}(K)$, and the price of the corresponding put option is $\text{put}_{0,t}(K)$.

Assumptions. I assume throughout the paper that there is no arbitrage, that there are no transaction costs or short sales constraints, and that there

	Assumption	VS	VIX	SVS	SVIX
A1	Arbitrary strikes	•	•	◦	•
A2	No dividends	•	•	◦	•
A3	Constant interest rate	•		•	
A4	Continuous trading	•		◦	
A5	No jumps	•			

Table 1: A bullet point (•) indicates that an assumption is required for results on variance swaps (VS), simple variance swaps (SVS), VIX, or SVIX. A hollow bullet point (◦) indicates that an assumption can be partially relaxed.

is no counterparty default. Table 1 summarizes other assumptions required in various sections of the paper.

A1 European puts and calls on the underlying asset can be traded with expiry date T and arbitrary strikes.

A2 The underlying asset does not pay dividends.

A3 The continuously-compounded interest rate is constant, at r.

A4 The underlying asset and the bond can be traded continuously in time.

A5 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.

All five assumptions are required to price and hedge variance swaps. Assumption A5 can be dropped entirely for simple variance swaps, which can be priced and hedged in the presence of jumps, and A1, A2 and A4 can be relaxed. Results 2 and 4 on the VIX and SVIX indexes only require A1–A2. Subsequent results are subject to assumptions discussed further below.

1 Variance swaps

A variance swap is an agreement to exchange

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

for some fixed "strike" $\widetilde{V}$ at time T. The market convention is to set $\widetilde{V}$ so that no money needs to change hands at initiation of the trade:

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

The following result, which is due to Carr and Madan (1998) and Demeterfi, Derman, Kamal, and Zou (1999), building on an idea of Neuberger (1994), shows how to price a variance swap in the $\Delta \to 0$ limit—that is, how to compute the expectation on the right-hand side of (2). From now on, $\widetilde{V}$ will always refer to the variance swap strike in this limiting case.

Result 1 (Pricing and hedging a variance swap in the $\Delta \to 0$ limit)**.** *Under Assumptions A1–A5, the strike on a variance swap is*

$$\widetilde{V} = 2e^{rT} \left\{ \int_0^{F_{0,T}} \frac{1}{K^2} \operatorname{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^{\infty} \frac{1}{K^2} \operatorname{call}_{0,T}(K)\, dK \right\}, \tag{3}$$

which has the interpretation

$$\widetilde{V} = \mathbb{E}^* \left[\int_0^T \sigma_t^2 \, dt \right]. \tag{4}$$

The variance swap can be hedged by holding

(i) a static position in $(2/K^2)\, dK$ puts expiring at time T with strike K, for each $K \leq F_{0,T}$,

(ii) a static position in $(2/K^2)\, dK$ calls expiring at time T with strike K, for each $K \geq F_{0,T}$, and

(iii) a dynamic position in $2(F_{0,t}/S_t - 1)/F_{0,T}$ units of the underlying asset at time t,

financed by borrowing.

Sketch proof. In the $\Delta \to 0$ limit, the expectation (2) converges to[1]

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

Neuberger (1994) observed that, by Itô's lemma and Assumption A5, $d \log S_t = (r - \frac{1}{2}\sigma_t^2) dt + \sigma_t \, dZ_t$ under the risk-neutral measure, so $(d \log S_t)^2 = \sigma_t^2 \, dt$, and

$$\begin{aligned} \widetilde{V} &= \mathbb{E}^* \left[\int_0^T \sigma_t^2 \, dt \right] \\ &= 2 \, \mathbb{E}^* \left[\int_0^T \frac{1}{S_t} \, dS_t - \int_0^T d \log S_t \right] \\ &= 2rT - 2 \, \mathbb{E}^* \log \frac{S_T}{S_0}. \end{aligned} \tag{5}$$

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 $R_T = S_T / S_0$. Carr and Madan (1998) and Demeterfi et al. (1999) then showed how to use the approach of Breeden and Litzenberger (1978) to find the price of this contract, P_{log}, in terms of the prices of European call and put options on the underlying asset:

$$P_{\mathsf{log}} \equiv e^{-rT} \, \mathbb{E}^* \log R_T = rTe^{-rT} - \int_0^{F_{0,T}} \frac{1}{K^2} \, \text{put}_{0,T}(K) \, dK - \int_{F_{0,T}}^{\infty} \frac{1}{K^2} \, \text{call}_{0,T}(K) \, dK. \tag{6}$$

Substituting (6) back into (5), we have the result. □

This result is often referred to as "model-free", since it allows the underlying asset's price to follow any Itô process. But Assumption A5 is strong, and it failed badly during the financial crisis of 2008–9. As Carr and Lee (2009) write, "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.

[1] Jarrow et al. (2010) provide a rigorous analysis.

1.1 The VIX index

Before turning to simple variance swaps, however, we pause to explore the properties of the VIX index, which is calculated based on option prices using an annualized and discretized version of (3), where the underlying asset is the S&P 500 index, and is generally interpreted as a measure of risk-neutral variance in the sense of (4). Working with the idealized version of VIX (i.e. not discretizing), we have

$$\mathrm{VIX}^2 \equiv \frac{2e^{rT}}{T}\left\{\int_0^{F_{0,T}} \frac{1}{K^2}\,\mathrm{put}_{0,T}(K)\,dK + \int_{F_{0,T}}^{\infty} \frac{1}{K^2}\,\mathrm{call}_{0,T}(K)\,dK\right\}. \tag{7}$$

Bollerslev, Tauchen and Zhou (2009) and Drechsler and Yaron (2011) use the VIX index (squared) to proxy for the risk-neutral expectation of the quadratic variation of log returns (the quantity on the right-hand side of equation (2)). In the the presence of jumps, however, the relationship between VIX and quadratic variation breaks down. To measure risk-neutral expected quadratic variation, one needs to observe the strikes on index variance swaps. Aït-Sahalia, Karaman and Mancini (2012) do so, obtaining over-the-counter data, and document a large gap between index variance swap strikes and VIX-type indices (squared) at all horizons: on the order of 2% in volatility units, compared to an average volatility level around 20%.

To summarize, if Assumptions A1–A5 held, VIX^2 would correspond to the strike on a variance swap, and would have the interpretation (4). But since jumps matter, VIX^2 does not correspond to the fair strike on a variance swap, $\widetilde{V}$; the replicating portfolio provided in Result 1 does not replicate the variance swap payoff; and neither $\widetilde{V}$ nor VIX^2 has the interpretation (4).

From now on, therefore, we will view (7) as a *definition*, not as a statement about variance swap pricing. The next result, which is valid even if the underlying asset's price can jump, shows that VIX measures the risk-neutral *entropy* of the simple return on the S&P 500 index. 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. Entropy makes appearances elsewhere in the finance literature:

see, for example, Alvarez and Jermann (2005), Backus, Chernov and Martin (2011), and Backus, Chernov and Zin (2013).

Result 2 (What does VIX measure?). *Under Assumptions A1–A2, VIX measures the risk-neutral entropy of the simple return:*

$$VIX^2 = \frac{2}{T}\,\mathrm{L}^*(R_T), \tag{8}$$

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

$$VIX^2 = \frac{1}{T}\operatorname{var}^* \log R_T \approx \frac{1}{T}\operatorname{var}^* R_T, \tag{9}$$

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,

$$VIX^2 = \sum_{n=2}^{\infty} \frac{2\kappa_n^*}{n!} = \kappa_2^* + \frac{\kappa_3^*}{3} + \frac{\kappa_4^*}{12} + \frac{\kappa_5^*}{60} + \cdots, \tag{10}$$

where $\kappa_n^* \equiv \frac{1}{T}\widetilde{\kappa}_n^*$*, and* $\widetilde{\kappa}_n^*$ *is the nth cumulant of* $\log R_T$*. (Thus* $\widetilde{\kappa}_1^* = \mathbb{E}^* \log R_T$*;* $\widetilde{\kappa}_2^* = \operatorname{var}^* \log R_T$*;* $\widetilde{\kappa}_3^*$ *is the risk-neutral skewness of* $\log R_T$*, multiplied by* $(\widetilde{\kappa}_2^*)^{3/2}$*;* $\widetilde{\kappa}_4^*$ *is the risk-neutral excess kurtosis, multiplied by* $(\widetilde{\kappa}_2^*)^2$*; and so on.)*

Proof. Equation (8) follows from the definition of VIX^2 and (6), together with the fact that $\mathbb{E}^* R_T = e^{rT}$. (The interest rate r is now interpreted as the continuously-compounded yield on a T-period zero-coupon bond.)

Equation (9) follows from (8) because $\log \mathbb{E}^* R_T = \mathbb{E}^* \log R_T + \frac{1}{2}\operatorname{var}^* \log R_T$ 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, $\operatorname{var}^* R_T = e^{2\mu}\left(e^{2\sigma^2} - e^{\sigma^2}\right) = e^{2\mu}\sigma^2 + O(\sigma^4) \approx \sigma^2 = \operatorname{var}^* \log R_T$.

To derive (10), we introduce the cumulant-generating function $\boldsymbol{\kappa}^*(\theta) = \log \mathbb{E}^*\left[e^{\theta \cdot \log R_T}\right]$, which 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 (10) 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, excess kurtosis, and so on, drop out in the lognormal case. $\square$

2 Simple variance swaps

Since variance swaps cannot be hedged at times of jumps, market participants have had to impose caps on their payoffs. These caps—which have become, since 2008, the market convention in index variance swaps as well as single-name variance swaps—limit the maximum possible payoff on a variance swap, but further complicate the pricing and interpretation of the contract. A fundamental problem with the definition of a conventional variance swap can be seen very easily: if the underlying asset—an individual stock, say—goes bankrupt, so that S_t hits zero at some point before expiry T, then the payoff (1) is *infinite*.

These considerations motivate the following definition. A simple variance swap is an agreement to exchange

$$\left(\frac{S_\Delta - S_0}{F_{0,0}}\right)^2 + \left(\frac{S_{2\Delta} - S_\Delta}{F_{0,\Delta}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{0,T-\Delta}}\right)^2 \tag{11}$$

for a pre-arranged strike V at time T. (Recall that $F_{0,t}$ is the forward price of the underlying asset to time t, which is known at time 0. If option prices are observable at all strikes, the forward price is the strike, K, at which $\text{call}_{0,t}(K) = \text{put}_{0,t}(K)$. In the special case in which the asset is non-dividend-paying, $F_{0,t}$ equals $S_0 e^{rt}$, so the denominators are geometrically increasing; and if the interest rate is also zero, then the denominators are all equal.)

Figure 7 in the appendix compares observed realized values of the payoffs (1) and (11) for variance swaps and simple variance swaps on the S&P 500 index, using $T = 1$ month and $\Delta = 1$ day. The two payoffs are generally similar, but the variance swap payoff (1) is larger if the S&P 500 index drops sharply over the trade horizon, and smaller if the S&P 500 index rises sharply.

As before, V is chosen 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 huge 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 importance.

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 the $\Delta \to 0$ limit. From now on, I write V for the fair strike on a simple variance swap in this limiting case, and write $V(\Delta)$ when the case of $\Delta > 0$ is considered.

The result allows for the underlying asset to pay dividends continuously at rate δS_t per unit time. (Dividends should be interpreted broadly: if the underlying asset is a foreign currency then δ corresponds to the foreign interest rate. Later in this section, and in the appendix, I consider other ways of modelling dividend payouts.) Given this assumption, $F_{0,t} = S_0 e^{(r-\delta)t}$.

Result 3 (Pricing and hedging a simple variance swap in the $\Delta \to 0$ limit). *Under Assumptions A1, A3 and A4, the strike on a simple variance swap is*

$$V = \frac{2e^{rT}}{F_{0,T}^2} \left\{ \int_0^{F_{0,T}} \text{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^{\infty} \text{call}_{0,T}(K)\, dK \right\}, \qquad (12)$$

and the payoff on a simple variance swap can be replicated by holding

(i) a static position in $(2/F_{0,T}^2)\, dK$ puts expiring at time T with strike K, for each $K \le F_{0,T}$,

(ii) a static position in $(2/F_{0,T}^2)\, dK$ calls expiring at time T with strike K, for each $K \ge F_{0,T}$, and

(iii) a dynamic position in $2e^{-\delta(T-t)}(1 - S_t/F_{0,t})/F_{0,T}$ units of the underlying asset at time t,

financed by borrowing.

Proof. The derivation of (12) divides into two steps.

Step 1. 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 it 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.,

$$\mathbb{E}\left[M_{(T)}\left\{\left(\frac{S_\Delta - S_0}{F_{0,0}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{0,T-\Delta}}\right)^2 - V\right\}\right] = 0. \tag{13}$$

We have

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

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

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

Substituting this into (13), we find that

$$V(\Delta) = \sum_{i=1}^{T/\Delta} \frac{e^{ri\Delta}}{F_{0,(i-1)\Delta}^2}\left[\Pi(i\Delta) - (2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta}\Pi((i-1)\Delta)\right]. \tag{14}$$

It remains to calculate $\Pi(t)$. But note that[2]

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

The right-hand side is the time-t payoff on an equal-weighted portfolio of European call options of all strikes, so a static no-arbitrage argument implies that

$$\Pi(t) = 2\int_0^\infty \mathrm{call}_{0,t}(K)\, dK. \tag{15}$$

[2]Darrell Duffie suggested this approach. A previous draft derived (15) via Breeden–Litzenberger (1978) logic and integration by parts: $\Pi(t) = \int_0^\infty K^2\, \mathrm{call}''_{0,t}(K)\, dK = 2\int_0^\infty \mathrm{call}_{0,t}(K)\, dK$. The latter approach is less neat, but has the advantage of being mechanical: it does not rely on a "trick".

To express $\Pi(t)$ in terms of out-of-the-money options, we can use the put-call parity relationship $\text{call}_{0,t}(K) = \text{put}_{0,t}(K) + e^{-rt}(F_{0,t} - K)$ to write this as

$$\Pi(t) = 2\int_0^{F_{0,t}} \text{put}_{0,t}(K)\,dK + 2\int_{F_{0,t}}^{\infty} \text{call}_{0,t}(K)\,dK + e^{-rt}F_{0,t}^2\,. \qquad (16)$$

Step 2. Observe that (14) can be rewritten

$$V(\Delta) = \sum_{i=1}^{T/\Delta}\left\{\frac{e^{ri\Delta}}{F_{0,(i-1)\Delta}^2}\left[P(i\Delta) - (2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta}P((i-1)\Delta)\right]\right\} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2,$$

where

$$P(t) \equiv 2\left\{\int_0^{F_{0,t}} \text{put}_{0,t}(K)\,dK + \int_{F_{0,t}}^{\infty} \text{call}_{0,t}(K)\,dK\right\}.$$

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

$$\frac{e^{rj\Delta}}{F_{0,(j-1)\Delta}^2} - \frac{e^{r(j+1)\Delta}}{F_{0,j\Delta}^2}(2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta} = \frac{e^{rj\Delta}}{F_{0,j\Delta}^2}\left(e^{(r-\delta)\Delta} - 1\right)^2.$$

(It may be helpful to note that definition (11) was originally found by viewing the normalizing constants $F_{0,j\Delta}$, for $j = 0, \ldots T/\Delta$, as arbitrary, and choosing them so that the above equation would hold.) We can therefore rewrite

$$V(\Delta) = \frac{e^{rT}}{F_{0,T-\Delta}^2}P(T) + \underbrace{\sum_{j=1}^{T/\Delta - 1}\frac{e^{rj\Delta}}{F_{0,j\Delta}^2}\left(e^{(r-\delta)\Delta} - 1\right)^2 P(j\Delta)}_{O(1/\Delta)\text{ terms of size }O(\Delta^2)} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2. \qquad (17)$$

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 $e^{rT}P(T)/F_{0,T}^2$, as required.

The above argument implicitly supplies the dynamic trading strategy that replicates the payoff on a simple variance swap. Appendix A describes the strategy in detail. □

The derivation of the pricing result (12) has two main components. The first is the exact expression (14), which applies for fixed $\Delta > 0$. It shows that 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, ..., T. But, correspondingly, the hedge 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, the second component shows that this concern is misplaced: by choosing forward prices as the normalizing weights in the definition (11), both the pricing formula (14) and the hedging portfolio simplify nicely in the limit as $\Delta \to 0$. In principle, we could have put any other constants known at time 0 in the denominators of the fractions in (11). 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, meaning that the hedge calls only for a single static portfolio of options expiring at time T, and equally weighted by strike.

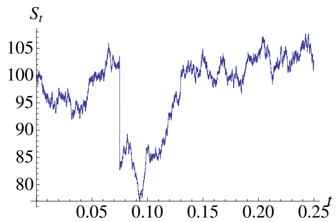
The dynamic position in the underlying can be thought of as a delta-hedge: if, say, the underlying's price at time t happens to exceed $F_{0,t} = S_0 e^{(r-\delta)t}$, then the replicating portfolio is short the underlying in order to offset the effects of increasing delta as calls go in-the-money and puts go increasingly out-of-the-money.

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.[3]

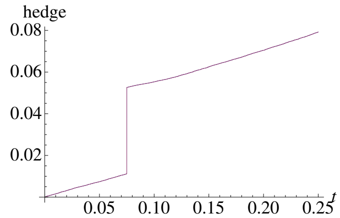
[3] Another recent innovation, the gamma swap, is closely related to a variance swap. At time T, a gamma swap pays

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

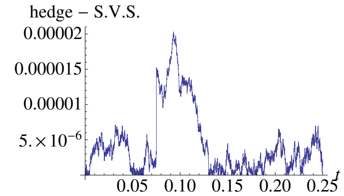
Under the Itô process assumption, gamma swaps can be replicated in a similar way to variance swaps; see Lee (2010a). In contrast with the hedge portfolio for a standard variance swap, which holds options with strikes K in amounts proportional to $1/K^2$, the hedge for a gamma swap holds portfolios of options with strikes K in amounts proportional to $1/K$. Simple variance swaps, which put equal weight on options of all strikes, are the next member of the sequence. But simple variance swaps are distinguished from the other two by the fact



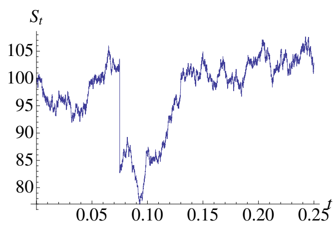
(a) Underlying price path



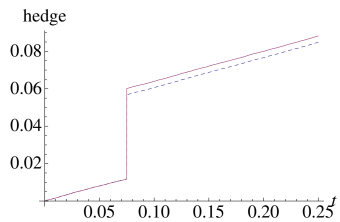
(b) SVS required payout and hedge portfolio performance



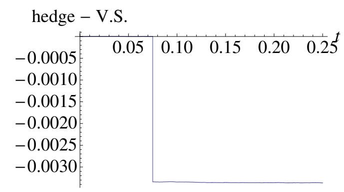
(c) SVS hedge error



(d) Underlying price path



(e) VS required payout and hedge portfolio performance



(f) VS hedge error

Figure 1: The payout and hedge error on a simple variance swap (SVS), and on a variance swap (VS), for a particular underlying price path.

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 simple variance swap (Figures 1a, 1b, and 1c) and for a 3-month variance swap (Figures 1d, 1e, and 1f). 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.[4] 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 (dashed line) substantially underperforms the required payout on the variance swap (solid line) due to the downward jump in the price of

that they can be priced and hedged in the presence of jumps.

[4]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. This example uses a discretization both in time and in the gap between strikes of options in the hedging portfolio. In the absence of this discretization, the hedge error on a simple variance swap would be *exactly* zero, as shown in Result 3.

the underlying. Panels 1c and 1f plot the difference between required payout and hedge portfolio performance in each case. The replicating portfolio for the variance swap suffers a large hedging error; in the case of the simple variance swap, the corresponding error is more than three orders of magnitude smaller.

Robustness. The above results rely on assumptions that are standard in the variance swap literature. It turns out, however, that simple variance swaps are robust to relaxing these assumptions in various ways.

First, what if sampling and trading occurs at discrete intervals $\Delta > 0$, rather than continuously? Equations (14) and (12) provide the fair strike on a simple variance swap in each case, and in Appendix B.1, I derive an analytic bound that shows that the gap between the two is extremely small if sampling is at daily, weekly or monthly intervals. I am not aware of any correspondingly general results in the variance swap literature—in fact, on the contrary, Broadie and Jain (2008) show, in the context of specific parametric models with jumps, that the fair strike on a variance swap can be significantly different depending whether sampling is continuous or discrete.

Second, what if deep-out-of-the-money options cannot be traded? This issue—which prevents perfect replication of variance swaps and simple variance swaps—cannot be entirely avoided so I show, in Appendix B.2, how to introduce an adjustment term to the contractually agreed payoff (11) on a simple variance swap if it is a significant concern. The resulting payoff can be replicated with the limited range of strikes that *is* tradable. I also show that in the case of S&P 500 index variance swaps, this adjustment term would equalled zero on every day in my sample period; in other words, even without the correction term, no problem would have materialized in sample. This indicates that jumps, rather than nontradability of deep-out-of-the-money options, were the fundamental problem for the index variance swap market during the recent crisis.

Third, one might worry about the effect of different dividend payout policies. But note that Result 3 continues to hold if the asset makes unanticipated dividend payouts. 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 (11): 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

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

In other words, the strategy perfectly replicates the desired payoff. This applies more generally: 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 Appendix B.3, I show how to modify the definition of the payoff (11) if, instead of paying dividends at rate δS_t, the asset pays dividends whose sizes and timing are known at time 0. Having done so, the results above go through almost unchanged.

The most challenging case is the fully general one, in which dividends are potentially anticipated, but of unknown size and timing. There is an elegant solution in this case too, assuming *dividend-adjusted* options can be traded: these are options on a claim to the underlying asset with dividends reinvested. Such options have recently started to trade on an over-the-counter basis, though they are relatively illiquid at present.[5] I will call the underlying with dividends reinvested the *dividend-adjusted underlying*. Then we can price and hedge a simple variance swap on the dividend-adjusted underlying directly from Result 3 simply by reinterpreting the inputs. The price S_t corresponds to the price of the dividend-adjusted underlying (so S_0 is the spot price of the underlying asset); the instantaneous dividend yield $\delta = 0$; $F_{0,t}$ is the forward price of the dividend-adjusted underlying, which equals $S_0 e^{rt}$ for all t by a static no-arbitrage argument; and $\text{put}_{0,T}(K)$ and $\text{call}_{0,T}(K)$ are the prices of dividend-adjusted options expiring at time T.

[5] I am grateful to Jack Busta for conversations on this point.

2.1 The SVIX index

From now on, the return R_T will always be the return on the S&P 500 index ("the market"). By analogy with VIX, we can define an index, SVIX, that is based on the annualized strike (12) of a simple variance swap:

$$\text{SVIX}^2 \equiv \frac{2e^{rT}}{T \cdot F_{0,T}^2} \left\{ \int_0^{F_{0,T}} \text{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^{\infty} \text{call}_{0,T}(K)\, dK \right\}. \tag{18}$$

In the remainder of the main body of the paper I assume that the underlying asset does not pay dividends, $\delta = 0$, to facilitate comparison with VIX. I also write $R_{f,T} = e^{rT}$ whenever it is convenient to do so.

Result 4 (What does SVIX measure?). *Under Assumptions A1–A2, SVIX measures the risk-neutral variance of the simple return:*

$$SVIX^2 = \frac{1}{T} \operatorname{var}^* \left(R_T / R_{f,T} \right). \tag{19}$$

Proof. (As in Result 2, the interest rate r is to be interpreted as the continuously-compounded yield on a T-period zero-coupon bond.) We have

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

(Since this equation follows from a static no-arbitrage argument, the link between SVIX and risk-neutral variance is unambiguously pinned down; this relationship applies whenever assumptions A1 and A2 hold, independent of whether or not the market is complete, for example.) From (16), this implies

$$\operatorname{var}^* R_T = \frac{2e^{rT}}{S_0^2} \left\{ \int_0^{F_{0,T}} \text{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^{\infty} \text{call}_{0,T}(K)\, dK \right\},$$

from which (19) follows. $\square$

If it seems surprising that the riskless rate enters equation (19), but not the corresponding equation (8) for the VIX index, then note that because entropy is invariant under scalings, $\mathrm{L}^*(R_T) = \mathrm{L}^*(R_T/R_{f,T})$.

Equation (19) has a nice graphical implication that is illustrated in Figure 2, which plots call and put prices against strike. Calls and puts have equal

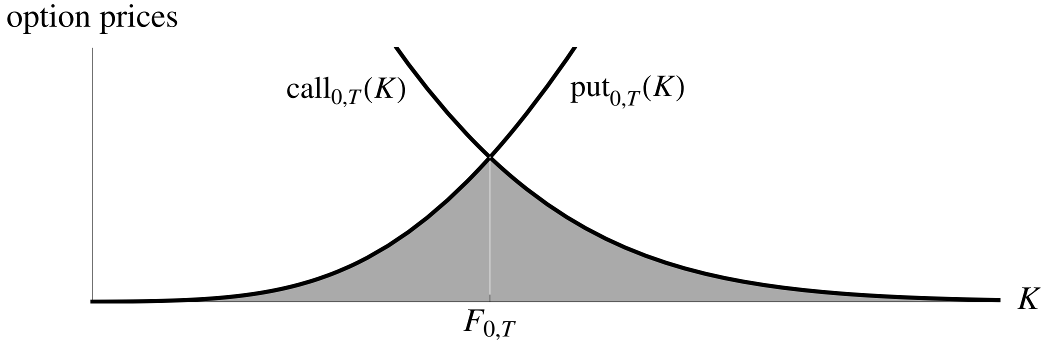


Figure 2: If the prices of call and put options expiring at time T are as shown, then the annualized risk-neutral variance of the underlying asset's simple return equals the shaded area under the curves multiplied by $2e^{rT}/(TS_0^2)$.

value when the strike equals the forward price, so the two lines intersect at $K = F_{0,T}$. The annualized risk-neutral variance is proportional to the shaded area under the two curves. SVIX is the square root of this quantity, so measures risk-neutral volatility.

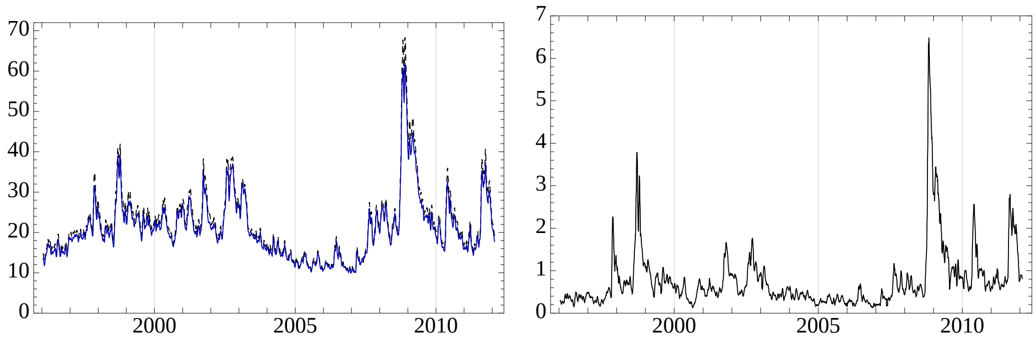
3 Comparing VIX and SVIX

If the VIX index measures entropy, and the SVIX index measures variance, which is a better measure of return variability? The answer is that both are of interest.[6] Entropy is more sensitive to the left tail of the return distribution, while variance is more sensitive to the right tail, as can be seen by comparing the entropy measure (7), which loads more strongly on out-of-the-money puts, with the variance measure (18), which loads equally on options of all strikes. I show below that the difference between VIX and SVIX is a measure of nonlognormality; and we will see in Section 4 that SVIX is a quantity that emerges naturally when connecting option prices to expected returns. Finally, note that the characterizations provided by Results 2 and 4 can be read in reverse as a way to calculate implied VIX and SVIX indices within equilibrium models: it is far easier to calculate risk-neutral entropy and variance than it

[6]Analogously, Hansen and Jagannathan (1991) study the variance of the stochastic discount factor, while Backus, Chernov and Martin (2012) and Backus, Chernov and Zin (2013) focus on its entropy. See the Online Appendix for further discussion.

is to compute option prices and then integrate over strikes.

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.



(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.

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 occur if returns R_T were lognormally distributed under the risk-neutral measure, as the next result will show, so loosely speaking, VIX minus SVIX is an index of nonlognormality. Unsurprisingly, therefore, the VIX–SVIX spread jumped in the recent crisis and at other times of market stress.

Result 5. *If the SDF M_T and return R_T are conditionally jointly lognormal, then $SVIX^2 = \frac{1}{T}(e^{\sigma_R^2 T} - 1)$ and $VIX^2 = \sigma_R^2$, where $\sigma_R^2 = \frac{1}{T}\operatorname{var}\log R_T$.*

Proof. In Appendix C. □

Since $e^{\sigma_R^2 T} - 1 > \sigma_R^2 T$, VIX is lower than SVIX in any conditionally lognormal model (and at the monthly or even annual horizon, the two would be almost equal). The opposite is true in the data, which 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.

Figure 8 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.

These results are consistent with the findings of Bollerslev and Todorov (2011a, 2011b).[7] Bollerslev and Todorov require identifying assumptions[8] 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.

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 (with, for example, time-varying conditional volatility). Figure 3b excludes that possibility; no model in which the market return and SDF are conditionally jointly lognormal is consistent with the data.

3.1 VIX and SVIX in equilibrium

Equation (10) has an implication that is surprising at first glance: if returns are more negatively skewed then, all else equal, VIX will be lower. 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.

[7] Several authors find similar results in parametric frameworks: for example, Bates (2000), Aït-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).

[8] 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).

We can do so by assuming that there is a marginal investor 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

$$\max_{\{w_i\}} \mathbb{E}\, u\left(\sum_i w_i R_T^{(i)}\right) \quad \text{subject to} \quad \sum_i w_i = 1. \tag{20}$$

The first-order conditions for this problem imply that $u'(R_T)/\,\mathbb{E}\,(R_T u'(R_T))$ is a stochastic discount factor. The proof of the next result exploits this fact.

Result 6 (Interpretation of VIX and SVIX)**.** *If there is a marginal investor with power utility and coefficient of relative risk aversion γ, then VIX and SVIX can be expressed in terms of the annualized real-world cumulants of* $\log R_T$*:*

$$\begin{aligned} \mathit{VIX}^2 &= \sum_{n=2}^{\infty} (-1)^n \alpha_n \kappa_n\,, \text{ where } \alpha_n = \frac{2\left[(\gamma-1)^n - \gamma^n + n\gamma^{n-1}\right]}{n!} && (21) \\ \mathit{SVIX}^2 &\approx \sum_{n=2}^{\infty} (-1)^n \beta_n \kappa_n\,, \text{ where } \beta_n = \frac{(\gamma-2)^n - 2(\gamma-1)^n + \gamma^n}{n!}. && (22) \end{aligned}$$

VIX and SVIX load equally on variance: $\alpha_2 = \beta_2 = 1$ for all γ. But VIX is more sensitively dependent on higher real-world cumulants: if $\gamma \geq 1$, then we have $\alpha_n > \beta_n \geq 0$ for $n > 2$. The approximation in (22) *is accurate so long as the horizon T is sufficiently short (equal to a year or less, say).*

Proof. In Appendix C. □

Result 6 should not be overemphasized, because it depends on the power utility assumption. But it confirms the basic intuition discussed above: although VIX is positively related to risk-neutral skewness (and higher odd risk-neutral cumulants), it is negatively related to their real-world counterparts, because $\alpha_n > 0$. The same is true for SVIX, because $\beta_n > 0$. But VIX is more sensitively dependent on higher cumulants than SVIX, because $\alpha_n > \beta_n$.

4 SVIX and the equity premium

SVIX can be connected to real-world (as opposed to risk-neutral) quantities in a rather general way. To see how, write M_T for the stochastic discount factor that prices time-T payoffs, and start from an identity:

$$\begin{aligned}\frac{\operatorname{var}^* R_T}{R_{f,T}} &= \mathbb{E}(M_T R_T^2) - R_{f,T} \\ &= \mathbb{E}\, R_T - R_{f,T} + \mathbb{E}(M_T R_T^2) - \mathbb{E}\, R_T \\ &= \mathbb{E}\, R_T - R_{f,T} + \underbrace{\operatorname{cov}(M_T R_T, R_T)}_{\leq 0 \text{ if NCC holds}}. \end{aligned} \tag{23}$$

This identity connects something that can be measured directly from SVIX (the risk-neutral variance) to something of interest (the equity premium) plus a covariance term. It motivates the following definition.

Definition 1 (Negative correlation condition)**.** *Given a gross return R_T and stochastic discount factor M_T, the* negative correlation condition (NCC) *holds if* $\operatorname{cov}(M_T R_T, R_T) \leq 0$.

The NCC is a convenient and flexible way to restrict the set of stochastic discount factors under consideration. It would, for example, fail badly in a risk-neutral economy, that is, if M_T were constant; empirically, though, M_T is extremely volatile (Hansen and Jagannathan (1991)). Roughly speaking, the NCC imposes two requirements: the stochastic discount factor must be volatile, and it must be negatively correlated with the return R_T.[9] In particular, the NCC holds if any of the following conditions holds.

B1 There is a one-period marginal investor who maximizes (20) above, and whose relative risk aversion $\gamma(x) \equiv -x u''(x)/u'(x)$, which need not be constant, satisfies $\gamma(x) \geq 1$.

B2 There is an intertemporal marginal investor with separable utility who holds the market, whose value function J can be defined recursively as

[9] If M_T and R_T are conditionally lognormal, the NCC is equivalent to the requirement that $-\operatorname{corr}(\log M_T, \log R_T)\,\sigma(\log M_T) \geq \sigma(\log R_T)$. For an alternative, and closely related, equivalence under lognormality, see condition B4 and Result 7, below.

a function of wealth W_0,

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

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

B3 There is an Epstein–Zin (1989) marginal investor who holds the market, and has a constant consumption-wealth ratio and risk aversion $\gamma \geq 1$.

B4 The SDF and market return are conditionally jointly lognormal, and the market's conditional Sharpe ratio exceeds its conditional volatility.

This list is far from exhaustive: one can, for example, adapt B1 or B2 to allow for situtations in which the marginal investor earns labor income, or holds bonds. Nonetheless, together these alternatives cover a reasonably wide range of models. Condition B1 is conceptually the simplest, but Cochrane (2011) has argued that it is crucial to allow for intertemporal considerations in such calculations. Condition B2 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. Condition B3 covers (a discrete-time version of) Wachter's (2011) time-varying disaster risk model. Condition B4 covers conditionally lognormal models—including, for example, those of Campbell and Cochrane (1999), Bansal and Yaron (2004), Bansal, Kiku, Shaliastovich and Yaron (2012) and Campbell, Giglio, Polk and Turley (2012)—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 (SVIX and the equity premium)**.** *If the NCC holds, then*

$$\mathbb{E}\,R_T - R_{f,T} \geq \frac{\operatorname{var}^* R_T}{R_{f,T}}. \tag{24}$$

If also assumptions A1–A2 hold, so that Result 4 applies, then SVIX provides a lower bound on the equity premium:

$$\frac{1}{T}\,\mathbb{E}\left[R_T - R_{f,T}\right] \geq R_{f,T} \cdot SVIX^2. \tag{25}$$

If any one of conditions B1–B4 holds, then the NCC holds; and if there is a marginal investor with log utility who holds the market, then the NCC and (24) *hold with equality.*

Proof. Inequality (24) follows from (23). Inequality (25) follows from (24) and Result 4. If there is a marginal investor with log utility who holds the market, then $M_T = 1/R_T$ is an SDF, and hence the NCC and inequality (24) hold with equality. It remains to show that any one of conditions B1–B4 implies that the NCC holds; I do so in Appendix C. □

Thus SVIX provides a direct measure of the forward-looking expected excess return on the market under the true, not the risk-neutral, probability distribution. This is a natural—perhaps *the* natural—measure of risk.

Result 7 provides a bound in the opposite direction from the Hansen–Jagannathan (1991) bound,

$$\frac{\mathbb{E}\, R_T - R_{f,T}}{\sigma(R_T)} \leq \frac{\sigma(M_T)}{\mathbb{E}\, M_T},$$

where $\sigma(\cdot)$ denotes conditional (real-world) standard deviation. The analogy between inequality (24) and the Hansen–Jagannathan bound can be brought out more strongly by rewriting (24) as

$$\frac{\mathbb{E}\, R_T - R_{f,T}}{R_{f,T}} \geq \left(\frac{\sigma^*(R_T)}{\mathbb{E}^*\, R_T} \right)^2 .$$

The Hansen–Jagannathan bound has the advantage of holding very generally, but the disadvantage that it relates two quantities that are not directly observable. As a result, time-series averages are typically used in practice to compute backward-looking Sharpe ratios as a proxy for the true measure of interest, the forward-looking Sharpe ratio. In contrast, Result 7 has the advantage of providing a directly observable bound on the equity premium, but the disadvantage that it relies on the NCC. The lower bound is of particular interest because, as we will see below, it was strikingly high during the crisis of 2008–9.

Inequality (25) is reminiscent of an approach taken by Merton (1980), based on the equation

$$\text{instantaneous risk premium} = \gamma\sigma^2 , \tag{26}$$

where γ is a measure of aggregate risk aversion, and σ^2 is the instantaneous variance of the market return. This relationship can be justified if there is a representative agent with constant relative risk aversion γ.

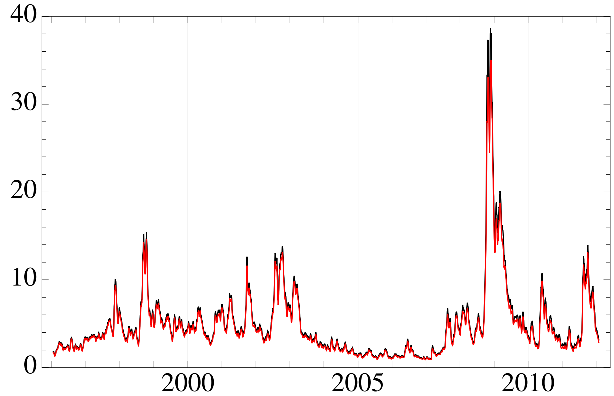
There are some important differences between the two approaches, however. First, Merton assumes that the market's price follows a diffusion, thereby ruling out the effects of skewness and of higher moments by construction.[10] In contrast, Result 7 makes no assumption about how prices evolve. Related 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. Once we move beyond diffusions, however, the appropriate generalization relates the risk premium to *risk-neutral* variance.

A second difference is that conditions B1 and B2 do not require that the marginal investor has *constant* relative risk aversion, only that the investor'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. One might imagine, for example, that in the CRRA case $\gamma(x) \equiv \gamma$, we could show that $\frac{1}{T}\,\mathbb{E}\left[R_T - R_{f,T}\right] = \gamma R_{f,T} \cdot \mathrm{SVIX}^2$. 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 $\frac{1}{T}\,\mathbb{E}\left[R_T - R_{f,T}\right] \geq \gamma R_{f,T} \cdot \mathrm{SVIX}^2$. 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.

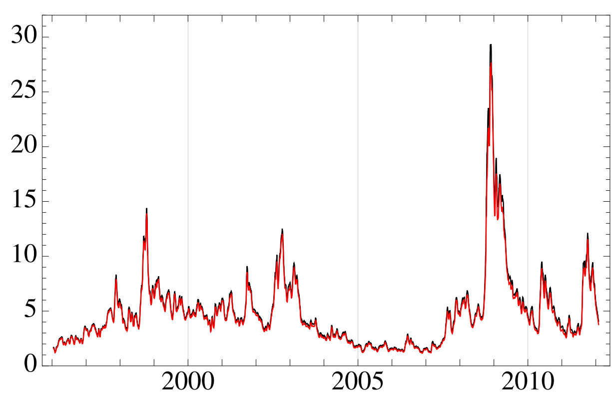
Third, Merton implements (26) using realized historical volatility rather than by exploiting option price data, though he notes that volatility measures could be calculated, in principle, "by 'inverting' the Black–Scholes option pricing formula". (Unfortunately—as he also notes—index options were not traded when he wrote the paper.) However, Black–Scholes implied volatility would only provide the correct measure of σ if we really lived in a Black–Scholes (1973) world in which prices followed geometric Brownian motions. The results of this paper show how to compute the right measure of variance in a more general environment.

Figure 4a plots the 10-day moving average of $R_{f,T} \cdot \mathrm{SVIX}^2$ measured in

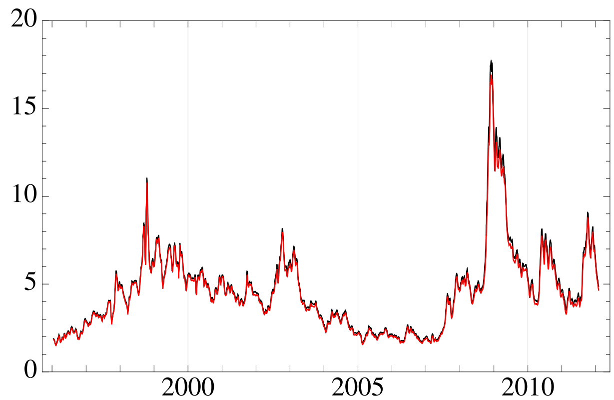
[10] Amongst others, Rubinstein (1973), Kraus and Litzenberger (1976), and Harvey and Siddique (2000) emphasize the importance of skewness in portfolio choice.



(a) 1 month



(b) 3 months



(c) 1 year

Figure 4: The lower bound on the annualized equity premium at different horizons (in %). The figures show 10-day moving averages. Mid prices in black; bid prices in red.

percentage points, at 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 5.00%—a number close to typical estimates of the unconditional equity premium. If either condition B1 or B2 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 considerable time-variation in the lower bound. During the "Great Moderation" 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. 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). The lower bound also has the appealing property that, by construction, it can never be less than zero. Most important, 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.

To address the concern that the lower bounds shown in Figure 4 are artificially high due to illiquidity in option markets, the figure also plots the lower bound calculated from the bid prices on options rather than mid prices. The resulting line, in red, is almost indistinguishable from the mid-price bound. The gap between the two is largest in November 2008, but the lower bound remains extremely high at all horizons.

Table 2 reports the mean, standard deviation, and various quantiles of the

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

Table 2: Mean, standard deviation, and quantiles of equity premium bounds at various horizons (annualized and measured in %).

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 VIX.

Consider, finally, a thought experiment. Suppose you find the lower bound on the equity premium in November 2008 implausibly high. What *trade* should you have done to implement this view? You should have sold a portfolio of options, namely an at-the-money-forward straddle and (equally weighted) out-of-the-money calls and puts. Such a position means that you end up short the market if the market rallies and long the market if the market sells off: essentially, you are taking a contrarian position, providing liquidity to the market. At the height of the credit crisis, extraordinarily high risk premia were available for investors who were prepared to take on this position.

Robustness. In practice, the idealized lower bound that emerges from the theory in the form of an integral (over option prices at all strikes) must be approximated by a sum (over option prices at observable strikes). In Appendix B.4, I show that the sums tend to underestimate the integrals; the key to this result is the well-known fact that option prices are convex functions of strike. The lower bound on the equity premium is therefore conservative: it would be even higher if option prices were observable at all strikes. Since the replacement of integrals by sums occurs throughout the literature, I also provide a corresponding result for the VIX index.

5 Conclusion

The theory of pricing and hedging of variance swaps depends on an assumption that prices follow diffusions, and hence cannot jump. When, in the recent crisis, prices *did* jump, the volatility derivatives market experienced turmoil. Individual stocks tend to jump more dramatically than indices, so the single-name variance swap market was affected particularly severely. It collapsed, and has not recovered.

Simple variance swaps are closely related to variance swaps. Unlike variance swaps, though, they can be priced and hedged even in the presence of jumps. I define an index, SVIX, that is based on the strike of a simple variance swap, much as the VIX index is based on the strike of a variance swap. But the link between SVIX and simple variance swaps holds even in the presence of jumps—unlike that between VIX and variance swaps.

The VIX and SVIX indices capture two different notions of the variability of the underlying return—entropy and variance, respectively—and the difference between the two, an index of non-lognormality, spikes up at times of stress. In the time series from January 1996 to January 2012, SVIX was always lower than VIX. This is direct evidence that we do not live in a lognormal world.

The paper concludes by linking the information in option prices to the market risk premium. I show that the SVIX index provides a lower bound on the forward-looking expected equity premium. This result relies on an assumption (the negative correlation condition) that holds in a range of standard models, and which captures the fact that the stochastic discount factor is volatile and negatively correlated with the market return. The bound does not require any stationarity or ergodicity assumptions, and it does not require statistical estimation of any parameters—so there are no in-sample/out-of-sample issues—because the lower bound is drawn directly from observable prices.

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 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%. In sharp contrast to the prevailing view in the literature, the SVIX index points to an equity premium that is extraordinarily volatile and that spiked dramatically at the height of the recent crisis.

6 References

Aït-Sahalia, Y., M. Karaman, and L. Mancini (2012), "The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis," working paper.

Aït-Sahalia, Y., Y. Wang, and F. Yared (2001), "Do Option Markets Correctly Price the Probabilities of Movement of the Underlying Asset?" *Journal of Econometrics* 102:67–110.

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.

Andersen, T. G., L. Benzoni, and J. Lund (2002), "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance* 57:3:1239–1284.

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

Backus, D. K., Chernov, M. and I. W. R. Martin (2011), "Disasters Implied by Equity Index Options," *Journal of Finance*, 66:6:1969–2012.

Backus, D. K., Chernov, M. and S. Zin (2013), "Sources of Entropy in Representative Agent Models," *Journal of Finance*, forthcoming.

Bakshi, G., N. Kapadia, and D. Madan (2003), "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *Review of Financial Studies* 16:1:101–143.

Bansal, R., D. Kiku, I. Shaliastovich, and A. Yaron (2012), "Volatility, the Macroeconomy, and Asset Prices," working paper.

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.

Bates, D. S. (2000), "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics* 94:181–238.

Bekaert, G., and E. Engstrom (2011), "Asset Return Dynamics under Bad Environment–Good Environment Fundamentals," working paper.

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

Bollerslev, T., and V. Todorov (2011a), "Tails, Fears, and Risk Premia," *Journal of Finance*, forthcoming.

Bollerslev, T., and V. Todorov (2011b), "Estimation of Jump Tails," *Econometrica*, forthcoming.

Bondarenko, O. (2007), "Variance Trading and Market Price of Variance Risk," working paper.

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

Broadie, M., M. Chernov, and M. Johannes (2007), "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance* 62:3:1453–1490.

Broadie, M. and A. Jain (2008), "The Effect of Jumps and Discrete Sampling on Volatility and Variance Swaps," *International Journal of Theoretical and Applied Finance*, 11:8:761–797.

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.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley, "An Intertemporal CAPM with Stochastic Volatility," working paper.

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.

Carr, P., and L. Wu (2003), "What Type of Process Underlies Options? A Simple Robust Test," *Journal of Finance* 58:6:2581–2610.

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.

Duffie, D. (2001), *Dynamic Asset Pricing Theory*, Princeton: Princeton University Press.

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.

Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance* 59:3:1367–1403.

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.

Harvey, C. R., and A. Siddique (2000), "Conditional Skewness in Asset Pricing Tests," *Journal of Finance* 55:3:1263–1295.

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.

Kraus, A., and R. H. Litzenberger (1976), "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance* 31:4:1085–1100.

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.
Neuberger, A. (2011), "Realized Skewness," working paper, Warwick Business School.
Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63:3–50.
Rubinstein, M. E. (1973), "The Fundamental Theorem of Parameter-Preference Security Valuation," *Journal of Financial and Quantitative Analysis* 8:1:61–69.
Summers, L. (1985), "On Economics and Finance," *Journal of Finance*, 40:3:633–635.
Wachter, J. (2011), "Can time-varying risk of rare disasters explain aggregate stock market volatility?" working paper, Wharton School of Business.

A Hedging a simple variance swap

The proof of Result 3 implicitly supplies the dynamic trading strategy that replicates the payoff on a simple variance swap. Tables 3 and 4 describe the strategy in detail. Each row of Table 3 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 (15); 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. One part of the proceeds is 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 3, but the general such position is indicated in Table 4, 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 3 adds up to the desired payoff,

$$\left(\frac{S_\Delta - S_0}{F_{0,0}}\right)^2 + \left(\frac{S_{2\Delta} - S_\Delta}{F_{0,\Delta}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{0,T-\Delta}}\right)^2 - V.$$

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 equation (14).

The replicating strategy simplifies nicely 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 4, this calls for a short position in the underlying asset of $2e^{-r(T-(j+1)\Delta)}S_{j\Delta}^2 e^{-\delta\Delta}/F_{0,j\Delta}^2$ in dollar terms at time $j\Delta$, that is, a short position of $2e^{-r(T-(j+1)\Delta)}S_{j\Delta}e^{-\delta\Delta}/F_{0,j\Delta}^2$ units of the underlying. In the limit as $\Delta \to 0$, holding $j\Delta = t$ constant, this equates to a short position of $2e^{-r(T-t)}S_t/F_{0,t}^2$ units of the underlying asset at time t.

The static position in options expiring at time T, shown in the penultimate line of Table 3, 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 (15). It is more natural, though, to use put-call parity to think of the position as a collection of calls with strikes above $F_{0,T}$ and puts with strikes below $F_{0,T}$, together with a long position in $2e^{-\delta(T-t)}/F_{0,T}$ 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 $2e^{-\delta(T-t)}/F_{0,T} - 2e^{-r(T-t)}S_t/F_{0,t}^2 = 2e^{-\delta(T-t)}(1 - S_t/F_{0,t})/F_{0,T}$ units of the asset and long out-of-the-money-forward calls and puts, all financed by borrowing.

B Robustness

This section collects the robustness results discussed in Sections 2 and 4.

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

Table 3: 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
$j\Delta$	$\frac{-(e^{(r-\delta)\Delta}-1)^2\Pi_{j\Delta}}{e^{r(T-j\Delta)}F^2_{0,j\Delta}}$	$\frac{(e^{(r-\delta)\Delta}-1)^2 S^2_{j\Delta}}{e^{r(T-j\Delta)}F^2_{0,j\Delta}}$		
B		$e^{-r(T-j\Delta)}\left[\frac{-S^2_{j\Delta}}{F^2_{0,(j-1)\Delta}} - \frac{S^2_{j\Delta}}{F^2_{0,j\Delta}}\right]$		$\frac{S^2_{j\Delta}}{F^2_{0,(j-1)\Delta}}+\frac{S^2_{j\Delta}}{F^2_{0,j\Delta}}$
U		$\frac{2S^2_{j\Delta}e^{-\delta\Delta}}{e^{r(T-(j+1)\Delta)}F^2_{0,j\Delta}}$	$\frac{-2S_{j\Delta}S_{(j+1)\Delta}}{e^{r(T-(j+1)\Delta)}F^2_{0,j\Delta}}$	
B			$\frac{2S_{j\Delta}S_{(j+1)\Delta}}{e^{r(T-(j+1)\Delta)}F^2_{0,j\Delta}}$	$\frac{-2S_{j\Delta}S_{(j+1)\Delta}}{F^2_{0,j\Delta}}$

Table 4: 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 $j\Delta$.

B.1 Pricing and hedging with $\Delta > 0$

The hedging strategy provided in Tables 3 and 4 perfectly replicates the desired payoff when $\Delta > 0$, but requires positions in options at all expiry dates $\Delta, \ldots, T-\Delta$. Discretizing the continuous-time strategy provided in the statement of Result 3 (which is exactly valid in the limit as $\Delta \to 0$) is equivalent to ignoring all such positions in options with intermediate expiry dates. The cashflows in these rows contribute a term of size $O(\Delta)$ at time 0, and terms of size $O(\Delta^2)$ at dates between 1 and $T-\Delta$. Thus the overall replication error is of size $O(\Delta)$, so the limiting strike is also an excellent approximation to the truth for sampling—and trading—intervals $\Delta > 0$. The next result makes this formal.

Result 8. *For $\Delta > 0$, the exact simple variance swap strike $V(\Delta)$, given by equation* (14)*, is very well approximated by V, given in equation* (12)*:*

$$|V(\Delta) - V| \leq \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2 (1+V) + \left|e^{2(r-\delta)\Delta} - 1\right| V\,. \tag{27}$$

The error term is tiny in practice: if $T = 1$, $r - \delta = 0.02$, $V = 0.05$, then the right-hand side of (27) *is less than 0.00001 with daily sampling ($\Delta = 1/252$),* *less than 0.00005 with weekly sampling* ($\Delta = 1/52$)*, and less than 0.0002 with monthly sampling* ($\Delta = 1/12$)*.*

Proof. Result 3 implies that for $j < T/\Delta$,

$$\frac{e^{rj\Delta}P(j\Delta)}{F_{0,j\Delta}^2} = \lim_{\Delta\to 0}\mathbb{E}^*\sum_{i=1}^{j}\left[\frac{S_{i\Delta}-S_{(i-1)\Delta}}{F_{0,(i-1)\Delta}}\right]^2 \le \lim_{\Delta\to 0}\mathbb{E}^*\sum_{i=1}^{T/\Delta}\left[\frac{S_{i\Delta}-S_{(i-1)\Delta}}{F_{0,(i-1)\Delta}}\right]^2 = \frac{e^{rT}P(T)}{F_{0,T}^2}.$$

Combining this observation with (17), we find that

$$\begin{aligned}\left|V(\Delta) - \frac{e^{rT}P(T)}{F_{0,T-\Delta}^2}\right| &= \sum_{j=1}^{T/\Delta-1}\left(e^{(r-\delta)\Delta}-1\right)^2\frac{e^{rj\Delta}P(j\Delta)}{F_{0,j\Delta}^2} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta}-1\right)^2 \\ &\le \frac{T}{\Delta}\left(e^{(r-\delta)\Delta}-1\right)^2\frac{e^{rT}P(T)}{F_{0,T}^2} + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta}-1\right)^2.\end{aligned}$$

Now, by definition of V, we have $|e^{rT}P(T)/F_{0,T-\Delta}^2 - V| = \left|e^{2(r-\delta)\Delta}-1\right|V$. Since $|V(\Delta)-V| \le |V(\Delta)-e^{rT}P(T)/F_{0,T-\Delta}^2| + |e^{rT}P(T)/F_{0,T-\Delta}^2 - V|$, by the triangle inequality, the result follows. □

B.2 Pricing and hedging when deep-out-of-the-money strikes are not tradable

Options that are sufficiently deep-out-of-the-money have prices so close to zero that they are not traded. Thus the idealized replicating portfolio, which comprises options of all strikes, is not attainable in practice. This issue affects both conventional variance swaps and simple variance swaps. Fortunately there is a practical solution to this problem. Suppose that, at time 0, options with strikes between A and B are tradable; the idealized scenario in which *all* strikes are tradable corresponds to $A = 0$, $B = \infty$. Then we can define the modified payoff

$$\left(\frac{S_\Delta - S_0}{F_{0,0}}\right)^2 + \left(\frac{S_{2\Delta}-S_\Delta}{F_{0,\Delta}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{0,T-\Delta}}\right)^2 - \phi(S_T), \qquad (28)$$

where the correction term $\phi(S_T)$ is zero unless the underlying asset's price happens to end up outside the original strike range (A, B):

$$\phi(S_T) = \begin{cases} \left(\frac{A - S_T}{F_{0,T-\Delta}}\right)^2 & \text{if } S_T < A. \\ 0 & \text{if } A \le S_T \le B. \\ \left(\frac{S_T - B}{F_{0,T-\Delta}}\right)^2 & \text{if } S_T > B. \end{cases}$$

The modified payoff (28) *can* be replicated without needing to trade options with strikes outside the range (A, B), by holding

(i) a static position in $2/F_{0,T}^2\, dK$ puts expiring at time T with strike K, for each $A < K \le F_{0,T}$,

(ii) a static position in $2/F_{0,T}^2\, dK$ calls expiring at time T with strike K, for each $F_{0,T} \le K < B$, and

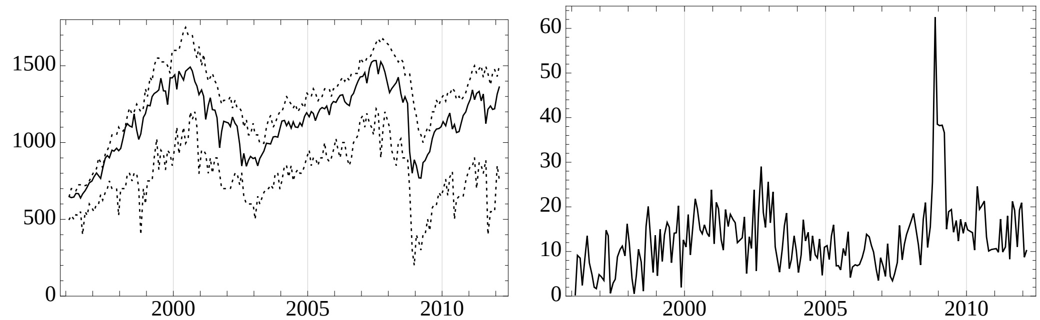
(iii) a dynamic position in $2e^{-\delta(T-t)}(1 - S_t/F_{0,t})/F_{0,T}$ units of the underlying asset at time t,

financed by borrowing. To see this, simply note that the payoff $\phi(S_T)$ is precisely the payoff on the "missing" options with strikes less than A and greater than B that are not included in the above position.

In the limit as $\Delta \to 0$, the fair strike that should be exchanged for the payoff (28) at time T is

$$\widehat{V} \equiv \frac{2e^{rT}}{F_{0,T}^2} \left\{ \int_A^{F_{0,T}} \text{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^B \text{call}_{0,T}(K)\, dK \right\}.$$

To explore how large the adjustment term $\phi(S_T)$ is in practice in the case of the S&P 500 index, I looked at every day in the sample on which *OptionMetrics* had data for options expiring in 30 days. On each such day, I recorded the lowest tradable strike (i.e. the strike of the most deep-out-of-the-money put option) and the highest tradable strike (i.e. the strike of the most deep-out-of-the-money call option), together with the subsequently realized level of the market at expiry time T.



(a) S_T, A, and B

(b) $\frac{\min\{S_T - A, B - S_T\}}{S_T}$

Figure 5: Left: Upper and lower strike boundaries, A and B (dotted lines) and subsequent realized level of the market at expiry, S_T (solid line). Right: The distance at expiry from the edge of the strike range, expressed as a percentage of the terminal level S_T.

The results are shown in Figure 5. Over the sample period, the underlying asset's price *never* ended up outside the range of tradable strikes. In other words, the correction term $\phi(S_T)$ was zero in every case: in Figure 5a, the value of S_T at expiry is within the range of strikes that were tradable at initiation on every day in sample. Figure 5b shows how far the underlying ended from the closer of the two boundaries, expressing the result as a percentage of S_T; the graph is always positive, reflecting the fact that the strike boundary was never crossed over the sample period. The spike in the figure occurred on 21 November, 2008, when the S&P 500 happened to close near the middle of the strike range that had prevailed 30 days previously; moreover, this occurred at a time when implied volatilities were very high, so that an extremely wide range of strikes had been traded. The low point in the figure occurred at the very beginning of the sample, on January 18, 1996, when the S&P 500 closed at 608.24. On that day, the highest strike tradable on options expiring in 30 days—on Saturday, February 17, 1996—was 650; in the event, the S&P 500 closed just two points lower, at 647.98, on Friday, February 16.

As is apparent from Figure 5a, the width of the range of tradable strikes has tended to increase over time. The mean value of the percentage distance to the edge of the strike range, as illustrated in Figure 5b, is 12.9%; the median value is 11.9%. In other words, on the median day in sample, the S&P 500 would have had to move a *further* 11.9% in the appropriate direction in order to exit the relevant range of tradable strikes.

B.3 Pricing and hedging if the underlying pays a known dividend

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 15.

In this section, it will be important to distinguish between $F_{0,t}$, the forward price of the dividend-paying asset to time t, and $\widetilde{F}_{0,t} \equiv S_0 e^{rt}$, 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

$$F_{0,t} = \begin{cases} S_0 e^{rt} & \text{if } t < k\Delta \\ \\ S_0 e^{rt} - D_{k\Delta} e^{r(t-k\Delta)} & \text{if } t \geq k\Delta \end{cases}$$

so $F_{0,t}$ and $\widetilde{F}_{0,t}$ coincide for times t before the payment of the dividend, but differ thereafter. It turns out that $\widetilde{F}_{0,t}$ 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

$$\left(\frac{S_\Delta - S_0}{\widetilde{F}_{0,0}}\right)^2 + \cdots + \left(\frac{S_{(k-1)\Delta} - S_{(k-2)\Delta}}{\widetilde{F}_{0,(k-2)\Delta}}\right)^2 + \left(\frac{S_{k\Delta} + D_{k\Delta} - S_{(k-1)\Delta}}{\widetilde{F}_{0,(k-1)\Delta}}\right)^2 + \\ + \left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{\widetilde{F}_{0,k\Delta}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{\widetilde{F}_{0,T-\Delta}}\right)^2 . \tag{29}$$

If the stock price happens to track the forward price at all points in time, then the payoff (29) 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 3 and 4 with δ set equal to zero (and replacing $F_{0,t}$ with $\widetilde{F}_{0,t}$ wherever it occurs in the tables). This replicating strategy generates the payoff (29) minus V, plus an extra payoff of $(D_{k\Delta}/\widetilde{F}_{0,(k-1)\Delta})^2 - 2D_{k\Delta}(S_{k\Delta} + D_{k\Delta})/\widetilde{F}^2_{0,(k-1)\Delta}$. To offset this extra payoff, two new positions are required: (i) a short position of $e^{-rT}(D_{k\Delta}/\widetilde{F}_{0,(k-1)\Delta})^2$ (measured in dollars) in bonds, and (ii) a long position of $2D_{k\Delta}e^{-r(T-k\Delta)}/\widetilde{F}^2_{0,(k-1)\Delta}$ 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

$$V = \frac{2e^{rT}}{\widetilde{F}^2_{0,T}} \left\{ \int_0^{F_{0,T}} \text{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^{\infty} \text{call}_{0,T}(K)\, dK \right\},$$

and that the replicating portfolio is equivalent to holding

(i) a static position of $2/\widetilde{F}^2_{0,T}$ puts expiring at time T with strike K, for each $K \leq F_{0,T}$,

(ii) a static position of $2/\widetilde{F}^2_{0,T}$ calls expiring at time T with strike K, for each $K \geq F_{0,T}$, and

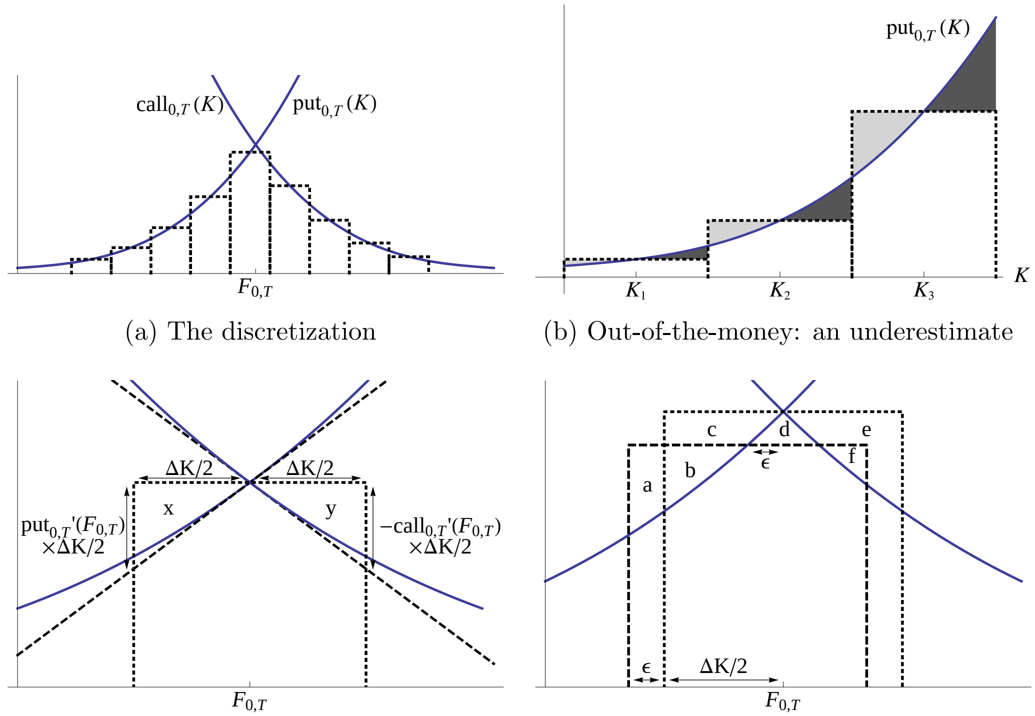
(iii) a dynamic position of $2(F_{0,t} - S_t)/(\widetilde{F}_{0,t}\widetilde{F}_{0,T})$ units of the underlying asset at time t,

financed by borrowing.

B.4 The effect of discretization by strike

The integrals that appear throughout the paper (and throughout the variance swap literature) are idealizations that cannot be implemented in practice because one only observes options with discrete strikes over a finite range. Write $\Omega_{0,T}(K)$ for the price of an out-of-the-money(-forward) option with strike K, that is,

$$\Omega_{0,T}(K) \equiv \begin{cases} \text{put}_{0,T}(K) & \text{if } K < F_{0,T} \\ \text{call}_{0,T}(K) & \text{if } K \geq F_{0,T} \end{cases};$$



(a) The discretization

(b) Out-of-the-money: an underestimate

(c) At-the-money: an overestimate

(d) The worst-case scenario in Lemma 1

Figure 6: The effect of discretization. Different panels use different scales.

write $K_1, \ldots, K_N$ for the strikes of observable options; write K_j for the strike that is nearest to the forward price $F_{0,T}$;[11] and define $\Delta K_i \equiv (K_{i+1} - K_{i-1})/2$. Then the idealized integral $\int_0^\infty \Omega_{0,T}(K)\, dK$ is replaced, in practice, by the observable sum $\sum_{i=1}^N \Omega_{0,T}(K_i)\, \Delta K_i$. (This is the CBOE's procedure in calculating VIX, and I follow it in this paper.) Figure 6a illustrates. The question is, how well does the sum approximate the integral?

The next result will show that there are two forces pushing in the direction of underestimation (of the integral by the sum) and one pushing in the direction of overestimation. But the latter effect is very minor in practice, so one should think of discretization as leading to underestimation of the integral.

Result 9 (The effect of discretization by strike)**.** *Discretizing by strike will*

[11]For simplicity, I assume that strikes are evenly spaced near-the-money, $K_{j+1} - K_j = K_j - K_{j-1}$. This is not essential, but it is almost always the case in practice and lets me economize slightly on notation.

tend to lead to an underestimate of the idealized value of SVIX^2, in that

$$\underbrace{\frac{2e^{rT}}{T\cdot F_{0,T}^2}\sum_{i=1}^{N}\Omega_{0,T}(K_i)\,\Delta K_i}_{\textit{discretization}} \leq \underbrace{\frac{2e^{rT}}{T\cdot F_{0,T}^2}\int_0^\infty \Omega_{0,T}(K)\,dK}_{\textit{idealized SVIX}^2} + \underbrace{\frac{(\Delta K_j)^2}{4T\cdot F_{0,T}^2}}_{\textit{very small}}.$$

The same is true for VIX^2, in that

$$\underbrace{\frac{2e^{rT}}{T}\sum_{i=1}^{N}\frac{\Omega_{0,T}(K_i)}{K_i^2}\,\Delta K_i}_{\textit{discretization}} \leq \underbrace{\frac{2e^{rT}}{T}\int_0^\infty \frac{\Omega_{0,T}(K)}{K^2}\,dK}_{\textit{idealized VIX}^2} + \underbrace{\frac{(\Delta K_j)^2}{4T\cdot F_{0,T}^2}}_{\textit{very small}}.$$

Proof. Non-observability of deep-out-of-the-money options obviously leads to an underestimation of SVIX^2.

Consider, first, the out-of-the-money puts with strikes $K_1,\ldots,K_{j-1}$. The situation is illustrated in Figure 6b: by convexity of $\text{put}_{0,T}(K)$, the light grey areas that are included (when they should be excluded) are smaller than the dark grey areas that are excluded (when they should be included). The same logic applies to the out-of-the-money calls with strikes $K_{j+1},K_{j+2},\ldots$. Thus the observable options—excluding the nearest-the-money option—will always underestimate the part of the integral which they are intended to approximate.

It remains to consider the nearest-the-money option with strike K_j, which alone can lead to an overestimate. Lemma 1, below, shows that the worst case is if the strike of the nearest-the-money option happens to be exactly *equal* to the forward price $F_{0,T}$, as in Figure 6c. For an upper bound on the overestimate in this case we must find an upper bound on the sum of the approximately triangular areas (x) and (y) that are shown in the figure. We can do so by replacing the curved lines in the figure by the (dashed) tangents to $\text{put}_{0,T}(K)$ and $\text{call}_{0,T}(K)$ at $K=F_{0,T}$. The areas of the resulting triangles provide the desired upper bound, by convexity of $\text{put}_{0,T}(K)$ and $\text{call}_{0,T}(K)$: we have

$$\text{area (x)} + \text{area (y)} \leq \frac{1}{2}\left(\frac{\Delta K}{2}\right)^2 \text{put}'_{0,T}(K) - \frac{1}{2}\left(\frac{\Delta K}{2}\right)^2 \text{call}'_{0,T}(K).$$

But, by put-call parity, $\text{put}'_{0,T}(K)-\text{call}'_{0,T}(K)=e^{-rT}$. Thus, the overestimate due to the at-the-money option is at most

$$\frac{1}{2}\left(\frac{\Delta K}{2}\right)^2 e^{-rT}.$$

Since the contributions from out-of-the-money and missing options led to underestimates, the overall overestimate is at most this amount. Finally, since the definition (18) scales the integral by $2e^{rT}/(TF_{0,T}^2)$, the result follows.

The proof of the result for VIX^2 follows exactly the same lines. The key observation is that $\text{call}_{0,T}(K)/K^2$ and $\text{put}_{0,T}(K)/K^2$ are convex, since they are each products of positive, monotonic, convex functions. $\square$

The maximal overestimate provided by this result is *extremely small*: for the S&P 500 index, the interval between strikes near-the-money is $\Delta K_j = 5$. If, say, the forward price of the S&P 500 index is $F_{0,T} = 1000$ and we are considering a monthly horizon, $T = 1/12$, then the discretization leads to an overestimate of SVIX^2 that is *at most* $7.5 \times 10^{-5} < 0.0001$. By comparison, the average level of SVIX^2 is on the order of 0.05, as shown in Table 2. Since the non-observability of deep-out-of-the-money options causes underestimation, there is therefore a strong presumption that the sum underestimates the integral.

It only remains to establish the following lemma, which is used in the proof of Result 9. The goal is to consider the largest possible overestimate that the option whose strike is nearest to the forward price, $F_{0,T}$, can contribute. Figure 6d illustrates. The dotted rectangle in the figure is the contribution if the strike happens to be *equal* to $F_{0,T}$; I will call this *Case 1.* The dashed rectangle is the contribution if the strike equals $F_{0,T} - \epsilon$, for some $\epsilon > 0$ (for concreteness—the case $\epsilon < 0$ is essentially identical); I will call this *Case 2.*

Lemma 1. *The option with strike closest to the forward overestimates most in the case in which its strike is equal to the forward.*

Proof. The overestimate in Case 1 is greater than that in Case 2 if

$$\text{area (b)}+\text{area (c)}+\text{area (e)}+\text{area (f)} \geq \text{area (a)}+\text{area (b)}+\text{area (f)}-\text{area (d)}$$

in Figure 6d, or equivalently,

$$\text{area (c)} + \text{area (d)} + \text{area (e)} \geq \text{area (a)}. \tag{30}$$

But, by convexity of $\text{put}_{0,T}(K)$,

$$\text{area (b)} + \text{area (c)} \geq \text{area (a)} + \text{area (b)},$$

from which (30) follows. An almost identical argument applies if $\epsilon < 0$. $\square$

C Proofs for Results 5, 6 and 7

Proof of Result 5. Let $R_T = e^{\mu_R T + \sigma_R \sqrt{T} Z_R - \sigma_R^2 T/2}$ and $M_T = e^{-rT + \sigma_M \sqrt{T} Z_M - \sigma_M^2 T/2}$, where Z_R and Z_M are Normal random variables with mean zero, variance one, and correlation ρ. Since $\mathbb{E}\, M_T R_T = 1$, we must have $\mu_R - r = -\rho\sigma_M\sigma_R$.

From (19), $\text{SVIX}^2 = \frac{e^{-2rT}}{T}\left[\mathbb{E}^*(R_T^2) - (\mathbb{E}^*\, R_T)^2\right] = \frac{1}{T}\left(e^{-rT}\,\mathbb{E}\, M_T R_T^2 - 1\right)$. Now, using the fact that $\mu_R - r = -\rho\sigma_M\sigma_R$, we have

$$\begin{aligned}
\mathbb{E}\, M_T R_T^2 &= \mathbb{E}\, e^{-rT + \sigma_M\sqrt{T}Z_M - \frac{1}{2}\sigma_M^2 T + 2\mu_R T + 2\sigma_R\sqrt{T}Z_R - \sigma_R^2 T} \\
&= e^{rT + \sigma_R^2 T}.
\end{aligned}$$

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

The calculation for VIX is slightly more complicated. Using (8), $\text{VIX}^2 = \frac{2}{T}\,\text{L}^*\, R_T = \frac{2}{T}\left[\log \mathbb{E}^*\, R_T - \mathbb{E}^* \log R_T\right] = \frac{2}{T}\left[rT - e^{rT}\,\mathbb{E}\, M_T \log R_T\right]$. Now,

$$\begin{aligned}
\mathbb{E}\left[M_T \log R_T\right] &= \mathbb{E}\left[(\mu_R T + \sigma_R\sqrt{T}Z_R - \frac{1}{2}\sigma_R^2 T)e^{-rT + \sigma_M\sqrt{T}Z_M - \sigma_M^2 T/2}\right] \\
&= (\mu_R - \frac{1}{2}\sigma_R^2)T \cdot e^{-rT} + \sigma_R\sqrt{T}e^{-rT - \sigma_M^2 T/2}\,\mathbb{E}\left[Z_R e^{\sigma_M\sqrt{T}Z_M}\right].
\end{aligned}$$

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

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

By Stein's lemma,

$$\begin{aligned}
\mathbb{E}\left[Z_M e^{\sigma_M\sqrt{T}Z_M}\right] &= \mathbb{E}\left[\sigma_M\sqrt{T}e^{\sigma_M\sqrt{T}Z_M}\right] \\
&= \sigma_M\sqrt{T}e^{\sigma_M^2 T/2}.
\end{aligned}$$

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

Proof of Result 6. 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 cumulant-generating function of the real-world cumulants $\widetilde{\kappa}_n$), and let $\boldsymbol{\kappa}^*(\theta)$ be the corresponding function calculated with respect to risk-neutral probabilities. The nth derivative of $\boldsymbol{\kappa}^*(\theta)$ at the origin equals the nth risk-neutral cumulant of $\log R_T$. In particular the first derivative at the origin equals the mean, $\mathbb{E}^* \log R_T = \boldsymbol{\kappa}^{*\prime}(0)$, so equations (8) and (19) relate VIX and SVIX to cumulants via the expressions

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

If $\boldsymbol{\kappa}^*(2) - 2\boldsymbol{\kappa}^*(1)$ is small (as I will now assume; this will be the case if the time horizon is a year or less, say), then $\text{SVIX}^2 \approx \frac{1}{T}[\boldsymbol{\kappa}^*(2) - 2\boldsymbol{\kappa}^*(1)]$, so

$$\text{SVIX}^2 \approx \sum_{n=2}^{\infty} \frac{(2^n - 2)\kappa_n^*}{n!} = \kappa_2^* + \kappa_3^* + \frac{7\kappa_4^*}{12} + \frac{\kappa_5^*}{4} + \cdots . \tag{31}$$

For comparison, Result 6 showed that

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

VIX and SVIX are equally sensitive to the risk-neutral variance of $\log R_T$, but SVIX is *more* sensitive to the higher *risk-neutral* cumulants.

To connect to the real-world cumulants, recall that we are assuming that the marginal investor has power utility. The SDF is therefore 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 $\boldsymbol{\kappa}^*(\theta) = \boldsymbol{\kappa}(\theta - \gamma) - \boldsymbol{\kappa}(-\gamma)$. So

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

which gives (21) and (22) after annualizing the cumulants by writing $\kappa_n \equiv \frac{1}{T}\widetilde{\kappa}_n$.

To show that, for $n > 2$ and $\gamma > 1$, we have $\alpha_n > \beta_n$, we must check that

$$2\left[(\gamma-1)^n - \gamma^n + n\gamma^{n-1}\right] > (\gamma-2)^n - 2(\gamma-1)^n + \gamma^n.$$

(The result is trivial if $\gamma = 1$.) Changing variables to $\widehat{\gamma} = \gamma - 1$, we must show that

$$2\left[\widehat{\gamma}^n - (1+\widehat{\gamma})^n + n(1+\widehat{\gamma})^{n-1}\right] > (\widehat{\gamma}-1)^n - 2\widehat{\gamma}^n + (1+\widehat{\gamma})^n.$$

This inequality relates two polynomials in $\widehat{\gamma}$. But notice that terms in $\widehat{\gamma}^n$ and $\widehat{\gamma}^{n-1}$ cancel on both sides. To establish the inequality, I will show that, for

$0 \leq m \leq n-2$, the coefficients on $\widehat{\gamma}^m$ are nonnegative, and larger on the left-hand side than on the right-hand side. Since $\widehat{\gamma} > 0$, this will establish the result.

Using the binomial theorem, the coefficient on $\widehat{\gamma}^m$ on the left-hand side is

$$2\left[-\binom{n}{m} + n\binom{n-1}{m}\right] = 2(n-m-1)\binom{n}{m},$$

where I use the fact that $n\binom{n-1}{m} = (n-m)\binom{n}{m}$. The coefficient on $\widehat{\gamma}^m$ in the right-hand side is $(-1)^{n-m}\binom{n}{m} + \binom{n}{m}$, which equals 0 if $n-m$ is odd, and $2\binom{n}{m}$ if $n-m$ is even. Thus, so long as $n-m-1 \geq 1$—that is, for $m \leq n-2$—the coefficient on the left-hand side is larger than that on the right-hand side whether $n-m$ is even or odd.

Since the coefficients on $\widehat{\gamma}^m$ on the right-hand side are strictly positive for even $n-m$, this also establishes that the right-hand side is greater than zero, that is, $\beta_n > 0$. $\square$

Conclusion of proof of Result 7. It remains to show that any one of conditions B1–B4 implies that the NCC holds. To do so under any of B1, B2, or B3, we will show that $M_T R_T$ is decreasing in R_T, from which the result follows. Condition B4 will be handled separately.

If B1 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 $u'(R_T) + R_T u''(R_T) = -u'(R_T)\left[\gamma(R_T) - 1\right] \leq 0$.

If B2 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, $J'[W_t] = u'(C_t)$, 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 marginal investor holds the market, the chosen portfolio weights w_i are such that $W_T = (W_0 - C_0)R_T$. That is, the SDF is proportional to $J'\left[(W_0 - C_0)R_T\right]$. We must therefore show that $R_T J'\left[(W_0 - C_0)R_T\right]$ is a decreasing function of R_T. This follows because its derivative with respect to R_T is $-J'\left[W_T\right]\left\{\Gamma(W_T) - 1\right\}$, which is less than or equal to zero because $\Gamma(x) \geq 1$.

If B3 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

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

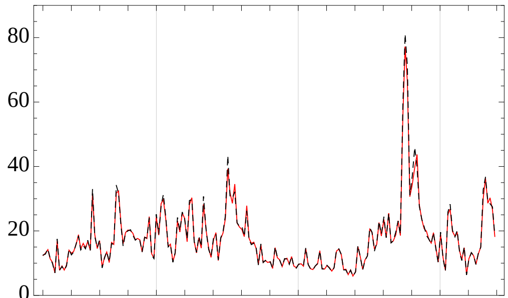
which gives the result.

If B4 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. Some straightforward algebra shows that $\mathbb{E}\, M_T R_T^2 \leq \mathbb{E}\, R_T$ if and only if $\lambda \geq \sigma_R$. $\square$

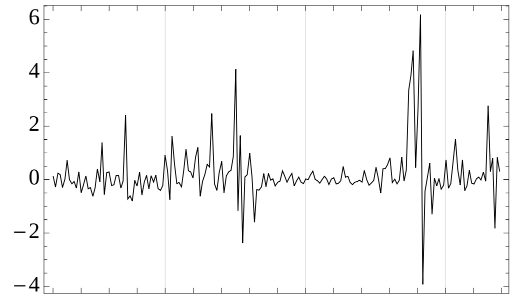
D Construction of SVIX

The data are from *OptionMetrics*, runnning 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, for each strike, I select the option—call or put—whose mid price is lower. 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 (18).

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; again, this is the procedure followed by the CBOE.

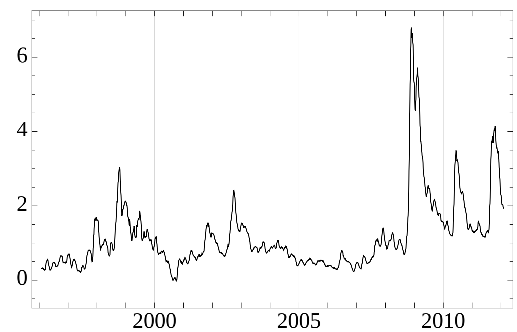


(a) Payoffs (1) and (11)

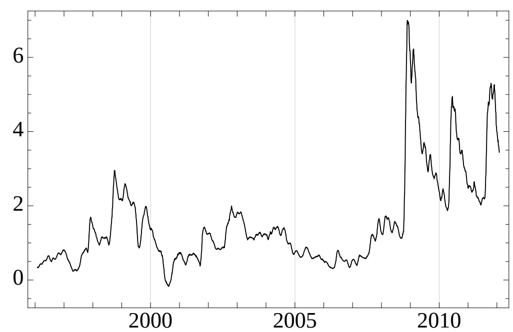


(b) Difference between payoffs (1) and (11)

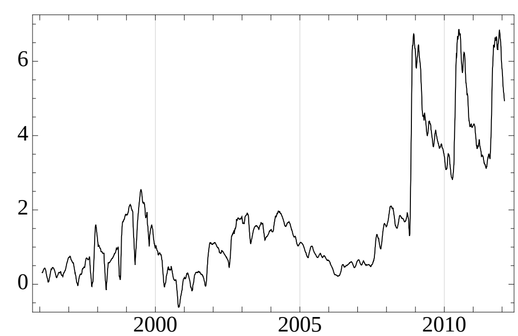
Figure 7: Left: Realized values of the payoffs (1) (dashed) and (11) (solid), expressed as annualized volatilities in percent. Right: The difference between the payoff (1) and the payoff (11), expressed as an annualized volatility in percent: $100 \times \sqrt{12} \times \left(\sqrt{\text{payoff (1)}} - \sqrt{\text{payoff (11)}}\right)$. In both panels, $T = 1$ month and $\Delta = 1$ day.



(a) 3 months



(b) 6 months



(c) 1 year

Figure 8: VIX–SVIX spread over different horizons.