WHAT IS THE EXPECTED RETURN ON THE MARKET?[*]

IAN MARTIN

I derive a lower bound on the equity premium in terms of a volatility index, SVIX, that can be calculated from index option prices. The bound implies that the equity premium is extremely volatile and that it rose above 20% at the height of the crisis in 2008. The time-series average of the lower bound is about 5%, suggesting that the bound may be approximately tight. I run predictive regressions and find that this hypothesis is not rejected by the data, so I use the SVIX index as a proxy for the equity premium and argue that the high equity premia available at times of stress largely reflect high expected returns over the very short run. I also provide a measure of the probability of a market crash, and introduce simple variance swaps, tradable contracts based on SVIX that are robust alternatives to variance swaps. *JEL Codes*: E44, G1.

I. INTRODUCTION

The expected excess return on the market, or equity premium, is one of the central quantities of finance and macroeconomics. Aside from its obvious intrinsic interest, the equity premium is a key determinant of the risk premium required for arbitrary assets in the capital asset pricing model (CAPM) and its descendants, and time variation in the equity premium lies at the heart of the literature on excess volatility.

The starting point of this article is an identity that relates the market's expected return to its risk-neutral variance. Under the weak assumption of no arbitrage, the latter can be measured unambiguously from index option prices. I call the associated volatility index SVIX and use the identity (coupled with a minimal assumption, the *negative correlation condition*, introduced in Section II) to derive a lower bound on the equity premium in terms of the SVIX index. The bound implies that the equity premium is extremely volatile, and that it rose above 21% at the height of the crisis in 2008. At horizons of less than a year, the equity premium fluctuates even more wildly: the lower bound on the monthly equity premium exceeded 4.5% (unannualized) in November 2008.

I go on to argue, more aggressively, that the lower bound appears empirically to be approximately tight, so that the SVIX index provides a direct measure of the equity premium. While it is now well understood that the equity premium is time-varying, this article deviates from the literature in its basic aim, which is to use theory to motivate a signal of expected returns that is based directly on asset prices. The distinctive features of the article, relative to the literature, are that (i) the predictor variable, $SVIX^2$, is motivated by asset pricing theory; (ii) no parameter estimation is required, so concerns over in-sample/out-of-sample fit do not arise; and (iii) since the $SVIX^2$ index is an asset price, I avoid the need to use infrequently updated accounting data. My approach therefore allows the equity premium to be measured in real time.

The $SVIX^2$ index can be interpreted as the equity premium perceived by an unconstrained rational investor with log utility who is fully invested in the market. This is a sensible benchmark even if there are many investors who are constrained and many investors who are irrational, and it makes for a natural comparison with survey evidence on investor expectations, as studied by Shiller (1987) and Ben-David, Graham, and Harvey (2013), among others. In particular, Greenwood and Shleifer (2014) emphasize the unsettling fact that the "expectations of returns" extracted from surveys are negatively correlated with subsequent realized returns. Greenwood and Shleifer also document the closely related fact that a range of survey measures of return expectations are negatively correlated with the leading predictor variables used in the literature to forecast expected returns. I show that the SVIX-based equity premium forecast is also negatively correlated with the survey measures of return expectations. But the SVIX
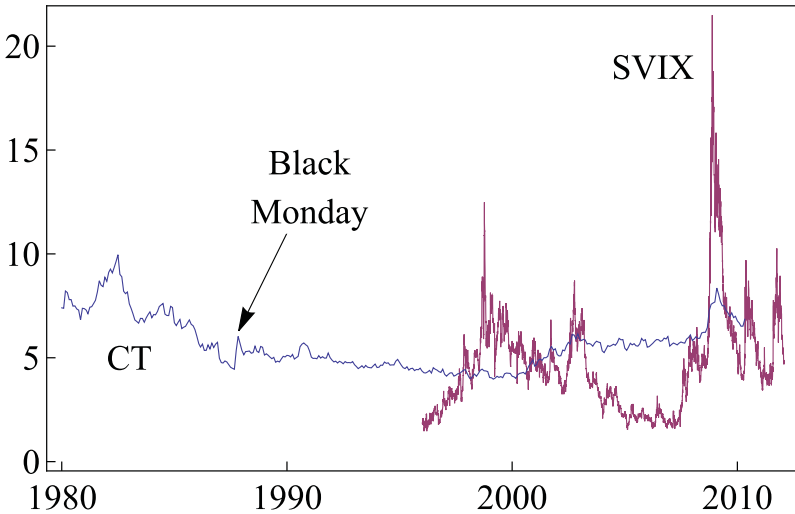
FIGURE I

Equity Premium Forecasts Based on Campbell and Thompson (2008) and on SVIX (Annual Horizon)

forecast is positively correlated with subsequent returns—a minimal requirement for a measure of rationally expected returns.

The view of the equity premium that emerges from the SVIX measure deviates in several interesting ways from the conventional view based on valuation ratios. Figure I plots the SVIX equity premium measure on the same axes as the smoothed earnings yield predictor of Campbell and Thompson (2008), whose work I take as representative of the vast predictability literature because their approach, like mine, avoids the in-sample/out-of-sample critique of Goyal and Welch (2008).[1] The figure illustrates the results of the article: I argue that the equity premium is more volatile, is more right-skewed, and fluctuates at a higher frequency than the literature has acknowledged.

I sharpen the distinction between the SVIX and valuation-ratio views of the world by focusing on two periods in which their predictions diverge. Valuation ratio-based measures of the equity premium were famously bearish throughout the late 1990s (and

---

1. Early papers in this literature include Keim and Stambaugh (1986), Campbell and Shiller (1988), and Fama and French (1988). A more recent paper that also argues for volatile discount rates is Kelly and Pruitt (2013). I thank John Campbell for sharing an updated version of the data set used in Campbell and Thompson (2008).

as noted by Ang and Bekaert 2007 and Goyal and Welch 2008, that prediction is partially responsible for the poor performance of valuation-ratio predictors in recent years); in contrast, the SVIX index suggests that at horizons up to one year, expected returns were high in the late 1990s. I suggest that this distinction reflects the fact that valuation ratios should be thought of as predictors of very long-run returns, whereas the SVIX index aims to measure short-run expected returns. The most striking divergence in predictions, however, occurs on one of the most dramatic days in stock market history, the great crash of October 1987, when option prices soared as the market collapsed.[2] On the valuation-ratio view of the world, the equity premium barely changed on Black Monday; on the SVIX view, it exploded.

II. Expected Returns and Risk-Neutral Variance

If we use asterisks to denote quantities calculated with risk-neutral probabilities, and M_T to denote the stochastic discount factor (SDF) that prices time T payoffs from the perspective of time t, then we can price any time T payoff X_T either via the SDF or by computing expectations with risk-neutral probabilities and discounting at the (gross) riskless rate, $R_{f,t}$, which is known at time t. The SDF notation,

$$(1) \qquad \text{time } t \text{ price of a claim to } X_T \text{ at time } T = \mathbb{E}_t(M_T X_T),$$

is commonly used in equilibrium models or, more generally, whenever there is an emphasis on the real-world distribution (whether from the subjective perspective of an agent within a model or from the "objective" perspective of the econometrician).

The risk-neutral notation,

$$(2) \qquad \text{time } t \text{ price of a claim to } X_T \text{ at time } T = \frac{1}{R_{f,t}} \mathbb{E}_t^* X_T,$$

is commonly used in derivative pricing or, more generally, whenever the underlying logic is that of no arbitrage. The choice of whether to use SDF or risk-neutral notation is largely a matter

2. Appendix Figure A.2, Panel A shows that implied volatility rose even more sharply on October 19, 1987, than it did in 2008–2009. As it turned out, the annualized return on the S&P 500 index was 81.2% over the month, and 23.2% over the year, following Black Monday.

of taste; I will tend to follow convention by using the risk-neutral notation when no-arbitrage logic is emphasized.

Equations (1) and (2) can be used to translate between the two notations; thus, for example, the conditional risk-neutral variance of a gross return R_T is

$$(3) \quad \mathrm{var}_t^* R_T = \mathbb{E}_t^* R_T^2 - \left(\mathbb{E}_t^* R_T\right)^2 = R_{f,t}\, \mathbb{E}_t\left(M_T R_T^2\right) - R_{f,t}^2.$$

Expected returns and risk-neutral variance are linked by the following identity:

$$\mathbb{E}_t R_T - R_{f,t} = \left[\, \mathbb{E}_t(M_T R_T^2) - R_{f,t} \right] - \left[\, \mathbb{E}_t\left(M_T R_T^2\right) - \mathbb{E}_t R_T \right]$$

$$(4) \qquad = \frac{1}{R_{f,t}}\, \mathrm{var}_t^* R_T - \mathrm{cov}_t(M_T R_T, R_T).$$

The first equality adds and subtracts $\mathbb{E}_t(M_T R_T^2)$; the second exploits equation (3) and the fact that $\mathbb{E}_t M_T R_T = 1$.

The identity (4) decomposes the asset's risk premium into two components. It applies to any asset return R_T, but in this article I focus on the case in which R_T is the return on the S&P 500 index. In this case the first component, risk-neutral variance, can be computed directly given time t prices of S&P 500 index options, as will be shown in Section IV. The second component is a covariance term that can be controlled: under a weak condition (discussed in detail in Section III), it is negative.

DEFINITION 1. Given a gross return R_T and stochastic discount factor M_T, the negative correlation condition (NCC) holds if $\mathrm{cov}_t\left(M_T R_T, R_T\right) \leqslant 0$.

Together, the identity (4) and the NCC imply the following inequality, from which the results of the article flow:

$$(5) \qquad \mathbb{E}_t R_T - R_{f,t} \geqslant \frac{1}{R_{f,t}}\, \mathrm{var}_t^* R_T.$$

This inequality can be compared to the bound of Hansen and Jagannathan (1991). The two inequalities place opposing bounds on the equity premium:

$$\frac{1}{R_{f,t}}\, \mathrm{var}_t^* R_T \leqslant \mathbb{E}_t R_T - R_{f,t} \leqslant R_{f,t} \cdot \sigma_t(M_T) \cdot \sigma_t(R_T),$$

where $\sigma_t(\cdot)$ denotes conditional (real-world) standard deviation. The left-hand inequality is (5). It has the advantage that it relates the unobservable equity premium to a directly observable quantity, risk-neutral variance; but the disadvantage that it requires the NCC to hold. In contrast, the right-hand inequality, the Hansen–Jagannathan bound, has the advantage of holding completely generally; but the disadvantage (noted by Hansen and Jagannathan) that it relates two quantities neither of which can be directly observed. Time-series averages must therefore be used as proxies for the true quantities of interest, forward-looking means and variances. This procedure requires assumptions about the stationarity and ergodicity of returns over appropriate sample periods and at the appropriate frequency. Such assumptions are not completely uncontroversial: see, for example, Malmendier and Nagel (2011).

The inequality (5) is reminiscent of the approach of Merton (1980) based on the equation

$$(6) \qquad \text{instantaneous risk premium} = \gamma \sigma^2,$$

where γ is a measure of aggregate risk aversion and σ^2 is the instantaneous variance of the market return, and of a closely related calculation carried out by Cochrane (2011, p. 1082).

There are some important differences between the two approaches, however. The first is that Merton assumes that the level of the stock index follows a geometric Brownian motion, thereby ruling out the effects of skewness and of higher cumulants by construction.[3] In contrast, we need no such assumption. 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 geometric Brownian motion, however, the appropriate generalization relates the risk premium to *risk-neutral* variance. As a bonus, this will have the considerable benefit that—unlike forward-looking real-world variance—forward-looking risk-neutral variance at time t can be directly and unambiguously computed from asset prices at time t, as I show in Section IV.

A second difference is that equation (6) requires that there is a representative agent with constant relative risk aversion γ.

---

3. Cochrane's calculation also implicitly makes this assumption; I show in Section VII.A that it is inconsistent with the data.

The NCC holds under considerably more general circumstances, as shown in Section III.

Third, Merton implements equation (6) using realized historical volatility rather than by exploiting option price data, though he notes that volatility measures can be calculated "by 'inverting' the Black–Scholes option pricing formula." However, Black–Scholes implied volatility would only provide the correct measure of σ if we really lived in the world of Black and Scholes (1973) in which prices follow geometric Brownian motions. The results of this article show how to compute the right measure of variance in a more general environment.

III. THE NEGATIVE CORRELATION CONDITION

This section examines the NCC more closely in the case in which R_T is the return on the market; it is independent of the rest of the article. I start by laying out various sufficient conditions for the NCC to hold. It is worth emphasizing that these conditions are not *necessary*: the NCC may hold even if none of the conditions below apply. The sufficient conditions cover many of the leading macro-finance models, including Campbell and Cochrane (1999), Bansal and Yaron (2004), Bansal et al. (2014), Campbell et al. (2016), Barro (2006), and Wachter (2013).[4]

The NCC is a convenient and flexible way to restrict the set of stochastic discount factors under consideration. It may be helpful to note that the NCC would fail badly in a risk-neutral economy—that is, if M_T were deterministic. We will need the SDF to be volatile, as is the case empirically (Hansen and Jagannathan 1991). We will also need the SDF to be negatively correlated with the return R_T; this will be the case for any asset that even roughly approximates the idealized notion of the market in economic models.[5]

The first example of this section indicates, in a conditionally log-normal setting, why the NCC is likely to hold in practice. It shows in particular that the NCC holds in several leading macro-finance models. All proofs for this section are in the Appendix.

4. In fact, I am not aware of any model that attempts to match the data quantitatively in which the NCC does not hold.

5. The NCC would fail for hedge assets (such as gold or, in recent years, U.S. Treasury bonds) whose returns tend to be high at times when the marginal value of wealth is high—that is, for assets whose returns are positively correlated with the SDF. Indeed, it may be possible to exploit this fact to derive *upper* bounds on the expected returns on such assets.

Example 1. Suppose that the SDF M_T and return R_T are conditionally log-normal and write $r_{f,t} = \log R_{f,t}$, $\mu_{R,t} = \log \mathbb{E}_t R_T$, and $\sigma_{R,t}^2 = \text{var}_t \log R_T$. Then the NCC is equivalent to the assumption that the conditional Sharpe ratio of the asset, $\lambda_t \equiv \frac{(\mu_{R,t} - r_{f,t})}{\sigma_{R,t}}$, exceeds its conditional volatility, $\sigma_{R,t}$.

The NCC therefore holds in any conditionally log-normal model in which the market's conditional Sharpe ratio is higher than its conditional volatility. Empirically, the Sharpe ratio of the market is on the order of 50% while its volatility is on the order of 16%, so it is unsurprising that this property holds in the calibrated models of Campbell and Cochrane (1999), Bansal and Yaron (2004), Bansal et al. (2014), and Campbell et al. (2016), among many others.

The special feature of the log-normal setting is that real-world volatility and risk-neutral volatility are one and the same thing.[6] So if an asset's Sharpe ratio is larger than its (real-world or risk-neutral) volatility, then its expected excess return is larger than its (real-world or risk-neutral) variance. That is, by (4), the NCC holds.

Unfortunately, the log-normality assumption is inconsistent with well-known properties of index option prices. The most direct way to see this is to note that equity index options exhibit a volatility smile: Black–Scholes implied volatility varies across strikes, holding option maturity constant. (See also Result 4 below.) This concern motivates the next example, which provides an interpretation of the NCC that is not dependent on a log-normality assumption.

Example 2. Suppose that there is an unconstrained investor who maximizes expected utility over next-period wealth, whose wealth is fully invested in the market, and whose relative risk aversion (which need not be constant) is at least one at all levels of wealth. Then the NCC holds for the market return. Moreover, if (but not only if) the investor has log utility, the covariance term in (4) is identically zero; then, the inequality (5) holds with equality, and $\mathbb{E}_t R_T - R_{f,t} = \frac{1}{R_{f,t}} \text{var}_t^* R_T$.

Example 2 does not require that the identity of the investor whose wealth is fully invested in the market should be fixed over time; thus it allows for the possibility that the portfolio holdings

---

6. More precisely, $\text{var}_t \log R_T = \text{var}_t^* \log R_T$ if M_T and R_T are conditionally jointly log-normal under the real-world measure.

and beliefs of (and constraints on) different investors are highly heterogeneous over time. Nor does it require that all investors are fully invested in the market, that all investors are unconstrained, or that all investors are rational. In view of the evidence presented by Greenwood and Shleifer (2014), this is an attractive feature. Under the interpretation of Example 2, the question answered by this article is this: what expected return must be perceived by an unconstrained investor with log utility who chooses to hold the market? This is a natural benchmark: there are many ways to be constrained, but only one way to be unconstrained. For reasons that will become clear in Sections V.B and VII.A, I prefer to interpret the data from the perspective of a log investor who holds the market, rather than the familiar representative investor who consumes aggregate consumption. Under this interpretation, my approach has nothing to say about—in particular, it does not resolve—the equity premium puzzle. In fact, on the contrary, the article documents yet another dimension on which existing equilibrium models fail to fit the data; see Section VII.A.

By focusing on a one-period investor, Example 2 abstracts from intertemporal issues and therefore from the presence of state variables that affect the value function. To the extent that we are interested in the behavior of long-lived utility-maximizing investors, we may want to allow for the fact that investment opportunities vary over time, as in the framework of Merton (1973). When will the NCC hold in (a discrete-time analog of) Merton's framework? Example 1 provided one answer to this question, but we can also frame sufficient conditions directly in terms of the properties of preferences and state variables, as in the next example (in which the driving random variables are normal, as in Example 1; this assumption will shortly be relaxed).

Example 3a. Suppose, in the notation of Cochrane (2005, pp. 166–167), that the SDF takes the form

$$M_T = \beta \frac{V_W(W_T, z_{1,T}, \ldots, z_{N,T})}{V_W(W_t, z_{1,t}, \ldots, z_{N,t})},$$

where W_T is the time T wealth of a risk-averse investor whose wealth is fully invested in the market, so that $W_T = (W_t - C_t)R_T$ (where C_t denotes the investor's time t consumption and R_T the return on the market); V_W is the investor's marginal value of wealth; and $z_{1,T}, \ldots, z_{N,T}$ are state variables, with signs chosen so that

V_W is weakly decreasing in each (just as it is weakly decreasing in W_T). Suppose also that

(i) Risk aversion is sufficiently high: $-\frac{WV_{WW}}{V_W} \geqslant 1$ at all levels of wealth W and all values of the state variables.

(ii) The market return, R_T, and state variables, $z_{1,T}, \ldots, z_{N,T}$, are increasing functions of conditionally normal random variables with (weakly) positive pairwise correlations.

Then the NCC holds for the market return.

Condition (i) imposes an assumption that risk aversion is at least 1, as in Example 2; again, risk aversion may be wealth- and state-dependent. Condition (ii) ensures that the movements of state variables do not undo the logic of Example 1. To get a feel for it, consider a model with a single state variable, the price-dividend ratio of the market (perhaps as a proxy for the equity premium, as in Campbell and Viceira 1999).[7] For consistency with the sign convention on the state variables, we need the marginal value of wealth to be weakly decreasing in the price-dividend ratio. It is intuitively plausible that the marginal value of wealth should indeed be high in times when valuation ratios are low; and this holds in Campbell and Viceira's setting, in the power utility case, if risk aversion is at least 1.[8] Then condition (ii) amounts to the (empirically extremely plausible) requirement that the correlation between the wealth of the representative investor and the market price-dividend ratio is positive. Equivalently, we need the return on the market and the market price-dividend ratio to be positively correlated. Again, this holds in Campbell and Viceira's calibration.

Example 3a assumes that the investor is fully invested in the market. Roll (1977) famously criticized empirical tests of the CAPM by pointing out that stock market indexes are imperfect proxies for the idealized notion of the market that may not fully capture risks associated with labor or other sources of income. Without denying the force of this observation, the implicit position taken is that although the S&P 500 index is not the sum

7. The price-dividend ratio is positive, so evidently cannot be normally distributed; this is why condition (ii) allows the state variables to be arbitrary increasing functions of normal random variables. For instance, we may want to assume that the log price-dividend ratio is conditionally normal, as Campbell and Viceira do.

8. Campbell and Viceira also allow for Epstein–Zin preferences, which I handle separately below.

total of all wealth, it *is* reasonable to ask, as a benchmark, what equity premium would be perceived by someone fully invested in the S&P 500. (In contrast, it would be much less reasonable to assume that some investor holds all of his wealth in gold in order to estimate the expected return on gold.)

Nonetheless, one may want to allow part of the investor's wealth to be held in assets other than the equity index. The next example allows for this possibility and generalizes in another direction by allowing the driving random variables to be non-normal.

Example 3b. Modify Example 3a by assuming that only a fraction α_t of wealth net of consumption is invested in the market (that is, in the equity index that is the focus of this article), with the remainder invested in some other asset or portfolio of assets that earns the gross return $R_T^{(i)}$:

$$W_T = \underbrace{\alpha_t(W_t - C_t)R_T}_{\text{market wealth},\, W_{M,T}} + \underbrace{(1 - \alpha_t)(W_t - C_t)R_T^{(i)}}_{\text{non-market wealth}}.$$

If the signs of state variables are chosen as in Example 3a, and if

(i) Risk aversion is sufficiently high: $-\frac{W V_{WW}}{V_W} \geqslant \frac{W_T}{W_{M,T}}$,

(ii) R_T, $R_T^{(i)}$, $z_{1,T}$, ..., $z_{N,T}$ are associated random variables,[9]

then the NCC holds for the market return.

Condition (i) shows that we can allow the investor's wealth to be less than fully invested in the market (for example, in bonds, housing, and human capital), as long as he cares more about the position he does have—that is, has higher risk aversion. If, say, at least a third of the investor's time T wealth is in the market, then the NCC holds as long as risk aversion is at least 3.

The next example handles models, such as Wachter (2013), that are neither conditionally log-normal nor feature investors with time-separable utility.

---

9. The concept of *associated* random variables (Esary, Proschan, and Walkup 1967) extends the concept of nonnegative correlation in a manner that can be extended to the multivariate setting. In particular, jointly normal random variables are associated if and only if they are nonnegatively correlated (Pitt 1982), and increasing functions of associated random variables are associated; thus Example 3a is a special case of Example 3b.

Example 4a. Suppose that there is a representative agent with Epstein–Zin (1989) preferences. If (i) risk aversion $\gamma \geqslant 1$ and elasticity of intertemporal substitution $\psi \geqslant 1$, and (ii) the market return R_T and wealth-consumption ratio $\frac{W_T}{C_T}$ are associated, then the NCC holds for the market return.

As special cases, condition (ii) would hold if, say, the log return $\log R_T$ and log wealth-consumption ratio $\log \frac{W_T}{C_T}$ are both normal and nonnegatively correlated; or if the elasticity of intertemporal substitution $\psi = 1$, since then the wealth-consumption ratio is constant (and hence, trivially, associated with the market return). This second case covers Wachter's (2013) model with time-varying disaster risk.

Example 4b. If there is a representative investor with Epstein–Zin–Weil preferences (Epstein and Zin 1989; Weil 1990), with risk aversion $\gamma = 1$ and arbitrary elasticity of intertemporal substitution then the NCC holds *with equality* for the market return. This case was considered (and not rejected) by Epstein and Zin (1991) and Hansen and Jagannathan (1991).

IV. Risk-Neutral Variance and the SVIX Index

We now turn to the question of measuring the risk-neutral variance that appears on the right-hand side of (5). The punchline will be that risk-neutral variance is uniquely pinned down by European option prices, by a static no-arbitrage argument. To streamline the exposition, I temporarily assume that the prices of European call and put options expiring at time T on the asset with return R_T are perfectly observable at all strikes K; this unrealistic assumption will be relaxed below.

Figure II plots a generic collection of time t prices of calls expiring at time T with strike K (written $\text{call}_{t,T}(K)$) and of puts expiring at time T with strike K (written $\text{put}_{t,T}(K)$). The figure illustrates two well-known facts that will be useful. First, call and put prices are convex functions of strike: any nonconvexity would provide a static arbitrage opportunity. This property will allow us to deal with the issue that option prices are only observable at a limited set of strikes. Second, the forward price of the underlying asset, $F_{t,T}$, which satisfies

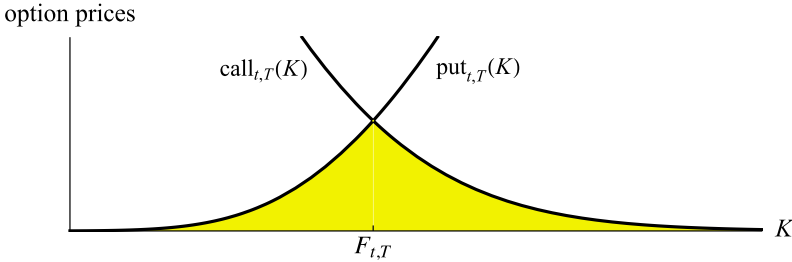$$(7) \qquad F_{t,T} = \mathbb{E}_t^* S_T,$$

option prices

FIGURE II

The Prices, at Time t, of Call and Put Options Expiring at Time T

can be determined by observing the strike at which call and put prices are equal, that is, $F_{t,T}$ is the unique solution x of the equation $\text{call}_{t,T}(x) = \text{put}_{t,T}(x)$. This fact follows from put-call parity; it means that the forward price can be backed out from time t option prices.

We want to measure $\frac{1}{R_{f,t}} \text{var}_t^* R_T$. I assume that the dividends earned between times t and T are known at time t and paid at time T,[10] so that

$$(8) \qquad \frac{1}{R_{f,t}} \text{var}_t^* R_T = \frac{1}{S_t^2} \left[\frac{1}{R_{f,t}} \mathbb{E}_t^* S_T^2 - \frac{1}{R_{f,t}} \left(\mathbb{E}_t^* S_T \right)^2 \right].$$

We can deal with the second term inside the square brackets using equation (7), so the challenge is to calculate $\frac{1}{R_{f,t}} \mathbb{E}_t^* S_T^2$. This is the price of the "squared contract"—that is, the price of a claim to S_T^2 paid at time T.

How can we price this contract, given put and call prices as illustrated in Figure II? Suppose we buy two call options with a strike of $K = 0.5$, two calls with a strike of $K = 1.5$, two calls with a strike of $K = 2.5$, two calls with a strike of $K = 3.5$, and so on, up to arbitrarily high strikes. The payoffs on the individual options are shown as dashed lines in Figure III, and the payoff on the portfolio of options is shown as a solid line. The idealized payoff S_T^2 is shown as a dotted line. The solid and dotted lines almost

---

10. If dividends are not known ahead of time, it is enough to assume that prices and dividends are (weakly) positively correlated, since then $\text{var}_t^* R_T \geqslant \text{var}_t^* \left(\frac{S_T}{S_t} \right)$, so that using $\frac{1}{R_{f,t}} \text{var}_t^* \left(\frac{S_T}{S_t} \right)$ instead of the ideal lower bound, $\frac{1}{R_{f,t}} \text{var}_t^* R_T$, is conservative.
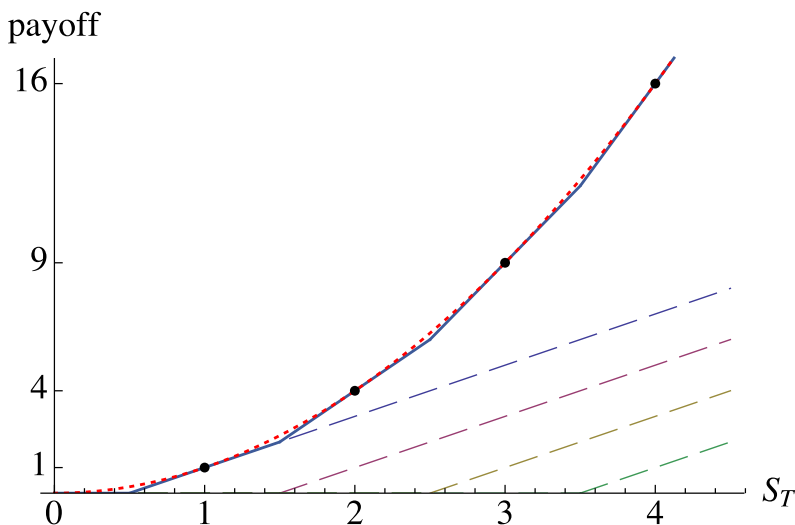
payoff

FIGURE III

Replicating the Squared Contract

The payoff S_T^2 (dotted line) and the payoff on a portfolio of options (solid line), consisting of two calls with strike $K = 0.5$, two calls with $K = 1.5$, two calls with $K = 2.5$, and so on. Individual option payoffs are indicated by dashed lines.

perfectly overlap, illustrating that the payoff on the portfolio is almost exactly S_T^2 (and it is *exactly* S_T^2 at integer values of S_T). Therefore, the price of the squared contract is approximately the price of the portfolio of options:

$$(9) \qquad \frac{1}{R_{f,t}} \mathbb{E}_t^* S_T^2 \approx 2 \sum_{K=0.5, 1.5, \dots} \text{call}_{t,T}(K).$$

To derive an exact expression, note that $x^2 = 2 \int_0^\infty \max\{0, x - K\} \, dK$ for any $x \geqslant 0$. Setting $x = S_T$, taking risk-neutral expectations, and multiplying by $\frac{1}{R_{f,t}}$,

$$\frac{1}{R_{f,t}} \mathbb{E}_t^* S_T^2 = 2 \int_0^\infty \frac{1}{R_{f,t}} \mathbb{E}_t^* \max\{0, S_T - K\} \, dK$$

$$(10) \qquad\qquad = 2 \int_0^\infty \text{call}_{t,T}(K) \, dK.$$

In practice, option prices are not observable at all strikes K, so we will need to approximate the idealized integral (10) by a sum along the lines of (9). To see how this will affect the results, notice that Figure III also demonstrates a subtler point: the option portfolio payoff is not just equal to the squared payoff at integers, it is *tangent* to it, so that the payoff on the portfolio of options very closely approximates *and is always less than or equal to* the ideal squared payoff. As a result, the sum over call prices in (9) will be slightly less than the integral over call prices in equation (10). This implies that the bounds presented are robust to the fact that option prices are not observable at all strikes: they would be even higher if all strikes were observable. Section IV.A expands on this point.

Combining equations (7), (8), and (10), we find that

$$\frac{1}{R_{f,t}} \operatorname{var}_t^* R_T = \frac{1}{S_t^2} \left[2 \int_0^\infty \operatorname{call}_{t,T}(K) \, dK - \frac{F_{t,T}^2}{R_{f,t}} \right].$$

Since deep-in-the-money call options are typically illiquid, it is convenient to split the range of integration into two and to replace in-the-money call prices with out-of-the-money put prices via the put-call parity formula $\operatorname{call}_{t,T}(K) = \operatorname{put}_{t,T}(K) + \frac{1}{R_{f,t}}(F_{t,T} - K)$, giving

$$\int_0^\infty \operatorname{call}_{t,T}(K) \, dK = \int_0^{F_{t,T}} \operatorname{put}_{t,T}(K) + \frac{1}{R_{f,t}}(F_{t,T} - K) \, dK$$

$$+ \int_{F_{t,T}}^\infty \operatorname{call}_{t,T}(K) \, dK$$

$$= \int_0^{F_{t,T}} \operatorname{put}_{t,T}(K) \, dK + \frac{F_{t,T}^2}{2R_{f,t}} + \int_{F_{t,T}}^\infty \operatorname{call}_{t,T}(K) \, dK.$$

Substituting back into equation (8), we find that

(11)
$$\frac{1}{R_{f,t}} \operatorname{var}_t^* R_T = \frac{2}{S_t^2} \left[\int_0^{F_{t,T}} \operatorname{put}_{t,T}(K) \, dK + \int_{F_{t,T}}^\infty \operatorname{call}_{t,T}(K) \, dK \right].$$

The expression in the square brackets is the shaded area shown in Figure II.

The right-hand side of equation (11) is strongly reminiscent of the definition of the VIX index, and indeed there are links that

will be explored in Section VII. To bring out the connection it will be helpful to define an index, $\text{SVIX}_{t \to T}$, via the formula

(12)

$$\text{SVIX}^2_{t \to T} = \frac{2}{(T-t)R_{f,t}S_t^2} \left[\int_0^{F_{t,T}} \text{put}_{t,T}(K)\,dK + \int_{F_{t,T}}^{\infty} \text{call}_{t,T}(K)\,dK \right].$$

The notation $\text{SVIX}_{t \to T}$ emphasizes that the SVIX index is calculated at time t based on the prices of options that mature at time T. It measures the annualized risk-neutral variance of the realized excess return from t to T: comparing equations (11) and (12), we see that

(13)
$$\text{SVIX}^2_{t \to T} = \frac{1}{T-t}\,\text{var}^*_t\left(\frac{R_T}{R_{f,t}}\right).$$

Inserting equation (11) into inequality (5), we have a lower bound on the expected excess return of any asset that obeys the NCC:

(14)

$$\mathbb{E}_t\, R_T - R_{f,t} \geqslant \frac{2}{S_t^2} \left[\int_0^{F_{t,T}} \text{put}_{t,T}(K)\,dK + \int_{F_{t,T}}^{\infty} \text{call}_{t,T}(K)\,dK \right],$$

or, in terms of the SVIX index,

(15)
$$\frac{1}{T-t}\left(\mathbb{E}_t\, R_T - R_{f,t}\right) \geqslant R_{f,t} \cdot \text{SVIX}^2_{t \to T}.$$

The bound will be applied in the case of the S&P 500; from now on, R_T always refers to the gross return on the S&P 500 index. I use option price data from OptionMetrics to construct a time series of the lower bound, at time horizons $T - t = 1, 2, 3, 6,$ and 12 months, from January 4, 1996, to January 31, 2012; the Appendix contains full details of the procedure. All results are annualized.

Panel A of Figure IV plots the lower bound, annualized and in percentage points, at the one-month horizon. Panels C and E of Figure IV repeat the exercise at three-month and one-year horizons. Table I reports the mean, standard deviation, and various quantiles of the distribution of the lower bound in the daily data for horizons between one month and one year.

The mean of the lower bound over the whole sample is 5.00% at the monthly horizon. This number is close to typical estimates
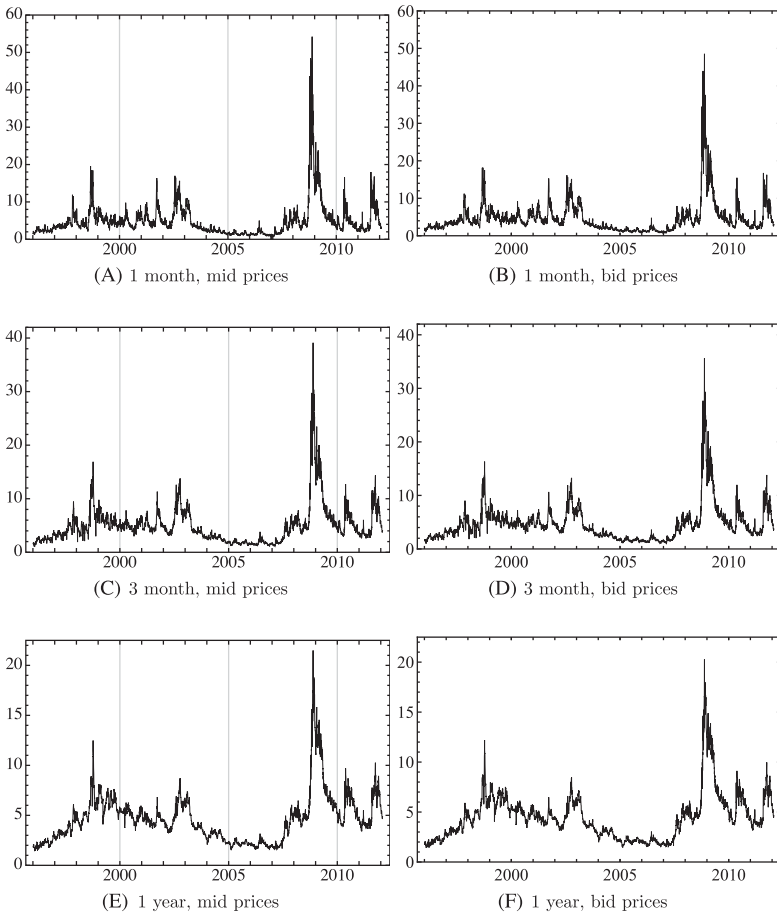
FIGURE IV

The Lower Bound on the Annualized Equity Premium at Different Horizons

The panels on the left use mid prices to calculate SVIX; those on the right use bid prices.

of the unconditional equity premium, which suggests that the bound may be fairly tight: that is, it seems that the inequality (14) may approximately hold with equality. Later I provide further tests of this possibility and develop some of its implications.

The time-series average of the lower bound is lower at the annual horizon than it is at the monthly horizon where the data quality is best (perhaps because of the existence of trades related

TABLE I

MEAN, STANDARD DEVIATION, SKEWNESS, AND QUANTILES OF THE LOWER BOUND ON THE EQUITY PREMIUM, $R_{f,t} \cdot SVIX^2_{t \to T}$, AT VARIOUS HORIZONS (ANNUALIZED AND MEASURED IN %)

Horizon	Mean	Std. dev.	Skew	Kurt	Min	1%	10%	25%	50%	75%	90%	99%	Max
1 mo	5.00	4.60	4.03	24.6	0.83	1.03	1.54	2.44	3.91	5.74	8.98	25.7	55.0
2 mo	5.00	3.99	3.37	17.5	1.01	1.20	1.65	2.61	4.11	5.91	8.54	23.5	46.1
3 mo	4.96	3.60	3.01	14.0	1.07	1.29	1.75	2.69	4.24	5.95	8.17	21.4	39.1
6 mo	4.89	2.97	2.37	9.13	1.30	1.53	1.95	2.88	4.39	6.00	7.69	16.9	29.0
1 yr	4.64	2.43	1.87	5.99	1.47	1.64	2.07	2.81	4.35	5.72	7.19	13.9	21.5

to VIX, which is itself a monthly index). It is likely that this reflects a less liquid market in one-year options, with a relatively smaller range of strikes traded, rather than an interesting economic phenomenon. I discuss this further in Section IV.A.

The lower bound is volatile, right-skewed, and fat-tailed. At the annual horizon the equity premium varies from a minimum of 1.22% to a maximum of 21.5% over my sample period. But variation at the one-year horizon masks even more dramatic variation over shorter horizons. The monthly lower bound averaged only 1.86% (annualized) during the Great Moderation years 2004–2006, but peaked at 55.0%—more than 10 standard deviations above the mean—in November 2008, at the height of the subprime crisis. Indeed, the lower bound hit peaks at all horizons 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.

Consider, finally, a thought experiment. Suppose you (an investor with preferences such that the NCC holds) 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 if the market sells off: 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 willing and able to take on this position.

IV.A. Robustness of the Lower Bound

Were option markets illiquid during the subprime crisis? One potential concern is that option markets may have been illiquid during periods of extreme stress. If so, one would expect to see a significant disparity between bounds based on mid-market option prices, such as those shown in the left panels of Figure IV, and bounds based on bid or offer prices, particularly in periods such as November 2008. Thus it is possible in principle that

the lower bounds would decrease significantly if bid prices were used. To address this issue, Panels B, D, and F of Figure IV show the corresponding bounds calculated from bid prices. Reassuringly, the results are very similar: the lower bound is high at all horizons whether mid or bid prices are used. Moreover, Appendix Figure A.1 shows that there was an increase in daily volume and open interest in S&P 500 index options over my sample period, and that the peaks in SVIX in 2008, 2010, and 2011 are associated with spikes in volume, rather than with the market drying up.

Option prices are only observable at a discrete range of strikes. Two issues arise when implementing the lower bound in practice. Fortunately, both issues mean that the numbers presented in this article are conservative: with perfect data, the lower bound would be even higher.

First, we do not observe option prices at all strikes K between 0 and ∞. This means that the range of integration in the integral we would ideally like to compute—the shaded area in Figure II—is truncated. Obviously, this will cause us to underestimate the integral in practice. This effect is likely to be strongest at the one-year horizon, because (in my data set) one-year options are less liquid than shorter-dated options.

Second, even within the range of observable strikes, prices are only available at a discrete set of strikes. Thus 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). What effect will this have? In the discussion of Figure III, I showed that the price of a particular portfolio of calls with a discrete set of strikes would very slightly underestimate the idealized measure, and hence be conservative. The general case, using out-of-the-money puts and calls, is handled in the Appendix, with the same conclusion.

V. SVIX AS PREDICTOR VARIABLE

The time-series average of the lower bound in recent data is approximately 5% in annualized terms, a number close to conventional estimates of the equity premium. Over the period 1951–2000, Fama and French (2002, Table IV) estimate the unconditional average equity premium to be 3.83% or 4.78%, based on dividend and earnings growth, respectively. It is therefore natural to wonder whether the lower bound might in fact be tight—that

TABLE II

COEFFICIENT ESTIMATES FOR REGRESSION (16)

Horizon	$\widehat{\alpha}$	Std. err.	$\widehat{\beta}$	Std. err.	R^2 (%)	R^2_{OS} (%)
1 mo	0.012	0.064	0.779	1.386	0.34	0.42
2 mo	−0.002	0.068	0.993	1.458	0.86	1.11
3 mo	−0.003	0.075	1.013	1.631	1.10	1.49
6 mo	−0.056	0.058	2.104	0.855	5.72	4.86
1 yr	−0.029	0.093	1.665	1.263	4.20	4.73

is, whether the lower bound illustrated in Figure IV is in fact a measure of the equity premium itself. We want to test the hypothesis that $\frac{1}{T-t}\left(\mathbb{E}_t\,R_T - R_{f,t}\right) = R_{f,t} \cdot \mathrm{SVIX}^2_{t\to T}$. Table II shows the results of regressions

$$(16) \qquad \frac{1}{T-t}\left(R_T - R_{f,t}\right) = \alpha + \beta \times R_{f,t} \cdot \mathrm{SVIX}^2_{t\to T} + \varepsilon_T,$$

together with robust standard errors that account for heteroskedasticity and overlapping observations (Hansen and Hodrick 1980). The null hypothesis that $\alpha = 0$ and $\beta = 1$ is not rejected at any horizon. The point estimates on β are close to 1 at all horizons, lending further support to the possibility that the lower bound is tight. This is encouraging because, as Goyal and Welch (2008) emphasize, this period is one in which conventional predictive regressions fare poorly.

One might worry that these results are entirely driven by the period in 2008 and 2009 in which volatility spiked and the stock market crashed before recovering strongly. To address this concern, Appendix Table A.1 shows the result of deleting all observations that overlap with the period August 1, 2008–July 31, 2009. Over horizons of one, two, and three months, deleting this period in fact increases the forecastability of returns by SVIX, reflecting the fact that the market continued to drop for a time after volatility spiked up in November 2008. On the other hand, the subsequent strong recovery of the market means that this was a period in which one-year options successfully predicted one-year returns, so by removing the crash from the sample, the forecasting power deteriorates at the one-year horizon.

We now have seen from two different angles that the lower bound (14) may be approximately tight: (i) as shown in Table I and Figure IV, the average level of the lower bound over my

sample is close to conventional estimates of the average equity premium; and (ii) Table II shows that the null hypothesis that $\alpha = 0$ and $\beta = 1$ in the forecasting regression (16) is not rejected at any horizon. These observations suggest that SVIX can be used as a measure of the equity premium without estimating any parameters—that is, imposing $\alpha = 0$, $\beta = 1$ in equation (16), so that

$$(17) \qquad \frac{1}{T-t} \left(\mathbb{E}_t R_T - R_{f,t} \right) = R_{f,t} \cdot \text{SVIX}^2_{t \to T} .$$

To assess the performance of the forecast (17), I follow Goyal and Welch (2008) in computing an out-of-sample R^2 measure

$$R^2_{OS} = 1 - \frac{\sum \varepsilon_t^2}{\sum v_t^2},$$

where ε_t is the error when SVIX (more precisely, $R_{f,t} \cdot \text{SVIX}^2_{t \to T}$) is used to forecast the equity premium and v_t is the error when the historical mean equity premium (computed on a rolling basis) is used to forecast the equity premium.[11]

The rightmost column of Table II reports the values of R^2_{OS} at each horizon. These out-of-sample R^2_{OS} values can be compared with corresponding numbers for forecasts based on valuation ratios, which are the subject of a vast literature.[12] Goyal and Welch (2008) consider return predictions in the form

$$\text{equity premium}_t = a_1 + a_2 \times \text{predictor variable}_t,$$

where a_1 and a_2 are constants estimated from the data, and argue that although conventional predictor variables perform reasonably well in-sample, they perform worse out-of-sample than does the rolling mean. Over their full sample (which runs from 1871 to 2005, with the first 20 years used to initialize estimates of a_1 and a_2, so that predictions start in 1891), the dividend-price ratio, dividend yield, earnings-price ratio, and book-to-market ratio have negative out-of-sample R^2s of -2.06%, -1.93%, -1.78%,

---

11. More detail on the construction of the rolling mean is provided in the Appendix.

12. Among many others, Campbell and Shiller (1988), Fama and French (1988), Lettau and Ludvigson (2001), and Cochrane (2008) make the case for predictability. Other authors, including Ang and Bekaert (2007), make the case against.

and -1.72%, respectively. The performance of these predictors is particularly poor over Goyal and Welch's recent sample (1976 to 2005), with R^2s of -15.14%, -20.79%, -5.98%, and -29.31%, respectively.[13]

Campbell and Thompson (2008) confirm Goyal and Welch's finding, and respond by suggesting that the coefficients a_1 and a_2 be fixed based on a priori considerations. Motivated by the Gordon growth model $\frac{D}{P} = R - G$ (where $\frac{D}{P}$ is the dividend-price ratio, R the expected return, and G expected dividend growth), Campbell and Thompson suggest making forecasts of the form

$$\text{equity premium}_t = \text{dividend-price ratio}_t + \text{dividend growth}_t$$
$$- \text{real interest rate}_t,$$

or, more generally,

$$\text{equity premium}_t = \text{valuation ratio}_t + \text{dividend growth}_t$$
$$(18) \qquad\qquad - \text{real interest rate}_t,$$

where in addition to the dividend-price ratio, Campbell and Thompson also consider earnings yields, smoothed earnings yields, and book-to-market as valuation ratios. As these forecasts are drawn directly from the data without requiring estimation of coefficients, they are a natural point of comparison for the forecast (17) suggested in this article.

Over the full sample, the out-of-sample R^2s corresponding to the forecasts (18) range from 0.24% (using book-to-market as the valuation ratio) to 0.52% (using smoothed earnings yield) in monthly data and from 1.85% (earnings yield) to 3.22% (smoothed earnings yield) in annual data.[14] The results are worse over Campbell and Thompson's most recent subsample, 1980–2005: in monthly data, R^2 ranges from -0.27% (book-to-market) to 0.03% (earnings yield). In annual data, the forecasts do even more poorly, each underperforming the historical mean, with R^2s ranging from -6.20% (book-to-market) to -0.47% (smoothed earnings yield).

---

13. Goyal and Welch show that the performance of an out-of-sample version of *cay* (Lettau and Ludvigson 2001) is similarly poor, with R^2 of -4.33% over the full sample and -12.39% over the recent sample.

14. Out-of-sample forecasts are from 1927 to 2005, or 1956 to 2005 when book-to-market is used.

In relative terms, therefore, the out-of-sample R^2s shown in Table II compare very favorably with the corresponding R^2s for predictions based on valuation ratios. But are they too small to be interesting in absolute terms? No. Ross (2005, pp. 54–57) and Campbell and Thompson (2008) point out that high R^2 statistics in predictive regressions translate into high attainable Sharpe ratios, for the simple reason that the predictions can be used to formulate a market-timing trading strategy; and if the predictions are very good, the strategy will perform extremely well. If Sharpe ratios above some level are too good to be true, then one should not expect to see R^2s from predictive regressions above some upper limit.

With this thought in mind, consider using risk-neutral variance in a contrarian market-timing strategy: each day, invest a fraction α_t in the S&P 500 index and the remaining fraction $1 - \alpha_t$ at the riskless rate, where α_t is chosen proportional to one-month SVIX^2 (scaled by the riskless rate, as on the right-hand side of equation (17)). The constant of proportionality has no effect on the strategy's Sharpe ratio, so I choose it such that the market-timing strategy's mean portfolio weight in the S&P 500 is 35%, with the remaining 65% in cash; the resulting median portfolio weight is 27% in the S&P 500, with 73% in cash. Figure V plots the cumulative return on an initial investment of $1 in this market-timing strategy and, for comparison, on strategies that invest in the short-term interest rate or in the S&P 500 index. In my sample period, the daily Sharpe ratio of the market is 1.35%, while the daily Sharpe ratio of the market-timing strategy is 1.97%; in other words, the out-of-sample R^2 of 0.42% reported in Table II is enough to deliver a 45% increase in Sharpe ratio for the market-timing strategy relative to the market itself. This exercise also illustrates the attractive feature that since risk-neutral variance is an asset price, it can be computed in daily data, or at even higher frequency, and thus permits high-frequency market-timing strategies to be considered.

As illustrated in Figure I, valuation ratios and SVIX tell qualitatively very different stories about the equity premium. First, option prices point toward a far more volatile equity premium than do valuation ratios. Second, SVIX is much less persistent than are valuation ratios, and so the SVIX predictor variable is less subject to Stambaugh (1999) bias. It is also noteworthy that SVIX forecasts a relatively high equity premium in the late 1990s. In this respect it diverges sharply from valuation ratio-based forecasts,
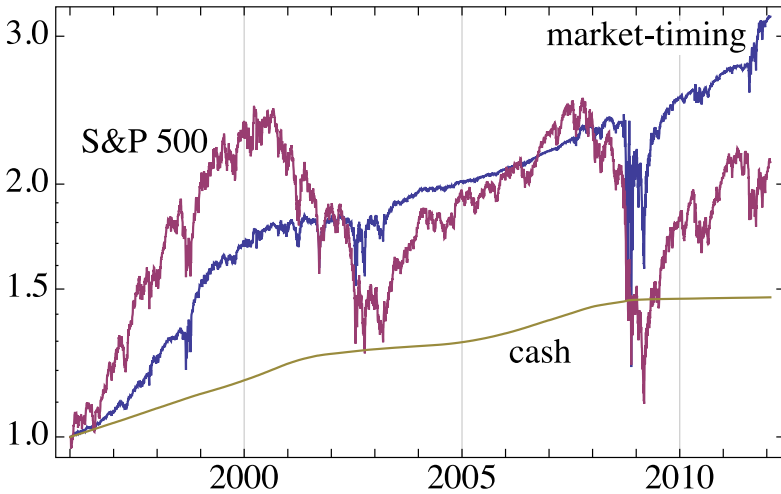
FIGURE V

Cumulative Returns on Three Strategies

Cumulative returns on \$1 invested in cash, in the S&P 500 index, and in a market-timing strategy whose allocation to the market at time t is proportional to $R_{f,t} \cdot \text{SVIX}^2_{t \to t+1\,\text{mo}}$ (log scale).

which predicted a low or even negative one-year equity premium at the time.

But perhaps the most striking aspect of Figure I is the behavior of the Campbell–Thompson predictor variable on Black Monday, October 19, 1987. This was by far the worst day in stock market history. The S&P 500 index dropped by over 20%—more than twice as far as on the second-worst day in history—and yet the valuation-ratio approach suggests that the equity premium barely responded. In sharp contrast, option prices exploded on Black Monday, implying that the equity premium was even higher than the peaks attained in November 2008.

V.A. The Term Structure of Equity Premia

Campbell and Shiller (1988) showed that any dividend-paying asset satisfies the approximate identity

$$d_t - p_t = \text{constant} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j \left(r_{t+1+j} - \Delta d_{t+1+j} \right),$$

which relates its log dividend yield $d_t - p_t$ to expectations of future log returns r_{t+1+j} and future log dividend growth Δd_{t+1+j}. Empirically, dividend growth is approximately unforecastable: to the extent that this is the case, we can absorb the terms $\mathbb{E}_t \, \Delta d_{t+1+j}$ into the constant, giving

$$(19) \qquad d_t - p_t = \text{constant} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j r_{t+1+j}.$$

This points a path toward reconciling the differing predictions of SVIX and valuation ratios. We can think of dividend yield as providing a measure of expected returns over the very long run. In contrast, the SVIX index measures expected returns over the short run.[15] The gap between the two is therefore informative about the gap between long-run and short-run expected returns. In the late 1990s, for example, $d_t - p_t$ was extremely low, indicating low expected long-run returns (Shiller 2000);[16] but Figure IV shows that SVIX, and hence expected *short-run* returns, were relatively *high* at that time.

We can also compare expected returns across shorter horizons. For example, Figure IV suggests that an unusually large fraction of the elevated one-year equity premium available in late 2008 was expected to materialize over the first few months of the 12-month period. To analyze this formally, define the annualized *forward equity premium* from T_1 to T_2 (calculated from the perspective of time t) by the formula

$$(20) \qquad \text{EP}_{T_1 \to T_2} \equiv \frac{1}{T_2 - T_1} \left(\log \frac{\mathbb{E}_t \, R_{t \to T_2}}{R_{f,t \to T_2}} - \log \frac{\mathbb{E}_t \, R_{t \to T_1}}{R_{f,t \to T_1}} \right),$$

15. It would be interesting to narrow the gap between long and short run by calculating SVIX indexes over the intermediate horizons that should be most relevant for macroeconomic aggregates such as investment. How do risk premia at, say, 5- or 10-year horizons behave? Data availability is a challenge here: long-dated options are relatively illiquid.

16. There is an important caveat. The discussion surrounding equation (19) follows much of the literature in blurring the distinction between expected arithmetic returns and the expected log returns that appear in the Campbell–Shiller log-linearization. Since $\mathbb{E}_t \, r_{t+1+j} = \log \mathbb{E}_t \, R_{t+1+j} - \frac{1}{2} \text{var}_t \, r_{t+1+j} - \sum_{n=3}^{\infty} \frac{\kappa_t^{(n)}(r_{t+1+j})}{n!}$, where $\kappa_t^{(n)}(r_{t+1+j})$ is the nth conditional cumulant of r_{t+1+j}, the gap between the two depends on the cumulants of log returns. So a low dividend yield may be associated with *high* expected arithmetic returns at times when log returns are highly volatile, right-skewed, or fat-tailed.

and the corresponding spot equity premium from time t to time T by

$$\mathrm{EP}_{t \to T} \equiv \frac{1}{T - t} \log \frac{\mathbb{E}_t \, R_{t \to T}}{R_{f, t \to T}}.$$

Using equation (17) to substitute out for $\mathbb{E}_t \, R_{t \to T_1}$ and $\mathbb{E}_t \, R_{t \to T_2}$ in the definition (20), we can write

$$\mathrm{EP}_{T_1 \to T_2} = \frac{1}{T_2 - T_1} \log \frac{1 + \mathrm{SVIX}_{t \to T_2}^2 (T_2 - t)}{1 + \mathrm{SVIX}_{t \to T_1}^2 (T_1 - t)}$$

and

$$\mathrm{EP}_{t \to T} = \frac{1}{T - t} \log \left(1 + \mathrm{SVIX}_{t \to T}^2 (T - t) \right).$$

(I have modified previous notation to accommodate the extra time dimension: for example, $R_{t \to T_2}$ is the simple return on the market from time t to time T_2 and $R_{f, t \to T_1}$ is the riskless return from time t to time T_1.)

The definition (20) is chosen so that for arbitrary $T_1, \ldots, T_N$, we have the decomposition

$$\mathrm{EP}_{t \to T_N} = \frac{T_1 - t}{T_N - t} \, \mathrm{EP}_{t \to T_1} + \frac{T_2 - T_1}{T_N - t} \, \mathrm{EP}_{T_1 \to T_2} + \cdots$$

(21)
$$+ \frac{T_N - T_{N-1}}{T_N - t} \, \mathrm{EP}_{T_{N-1} \to T_N},$$

which expresses the long-horizon equity premium $\mathrm{EP}_{t \to T_N}$ as a weighted average of forward equity premia, exactly analogous to the relationship between spot and forward bond yields.

Figure VI shows how the annual equity premium decomposes into a one-month spot premium plus forward premia from 1 to 2, 2 to 3, 3 to 6, and 6 to 12 months. The figure stacks the unannualized forward premia—terms of the form $\frac{(T_n - T_{n-1})}{(T_N - t)} \mathrm{EP}_{T_{n-1} \to T_n}$—which add up to the annual equity premium, as shown in equation (21). For example, on any given date t, the gap between the top two lines represents the contribution of the unannualized 6-month-6-month-forward equity premium, $\frac{1}{2} \mathrm{EP}_{t+6\mathrm{mo} \to t+12\mathrm{mo}}$, to the annual equity premium, $\mathrm{EP}_{t \to t+12\mathrm{mo}}$.

In normal times, the 6-month-6-month-forward equity premium contributes about half of the annual equity premium, as
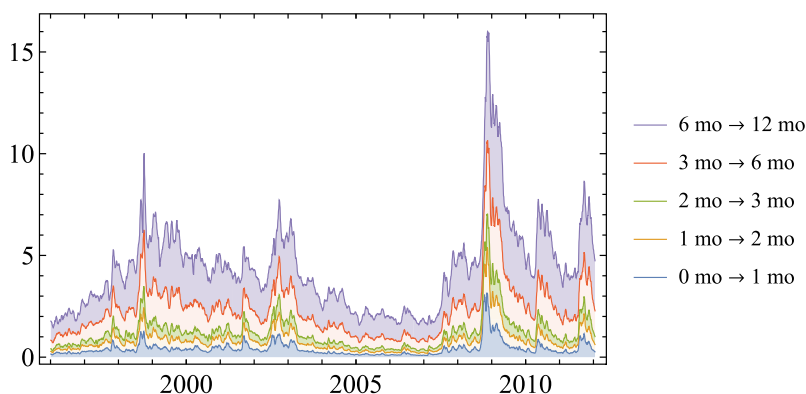
FIGURE VI

The Term Structure of Equity Premia (10-Day Moving Average)

might have been expected. More interesting, the figure shows that at times of stress, much of the annual equity premium is compressed into the first few months. For example, about a third of the equity premium over the year from November 2008 to November 2009 can be attributed to the (unannualized) equity premium over the two months from November 2008 to January 2009.

V.B. *Expectations of Returns and Expected Returns*

The view of the equity premium proposed above can usefully be compared with the expectations reported in surveys of market participants, as studied by Shiller (1987), Ben-David, Graham, and Harvey (2013), and others. In particular, Greenwood and Shleifer (2014) emphasize that survey-based return expectations are negatively correlated with expected return forecasts based on conventional predictor variables. We will now see that this is also true when SVIX is used as a predictive variable.

Figure VII shows four of the survey measures considered by Greenwood and Shleifer: the Graham–Harvey chief financial officer surveys, the Gallup investor survey measure, the American Association of Individual Investors (AAII) survey, and Robert Shiller's investor survey. The Graham–Harvey survey is based on the expectations of market returns reported by the chief financial officers of major U.S. corporations; this survey can be compared
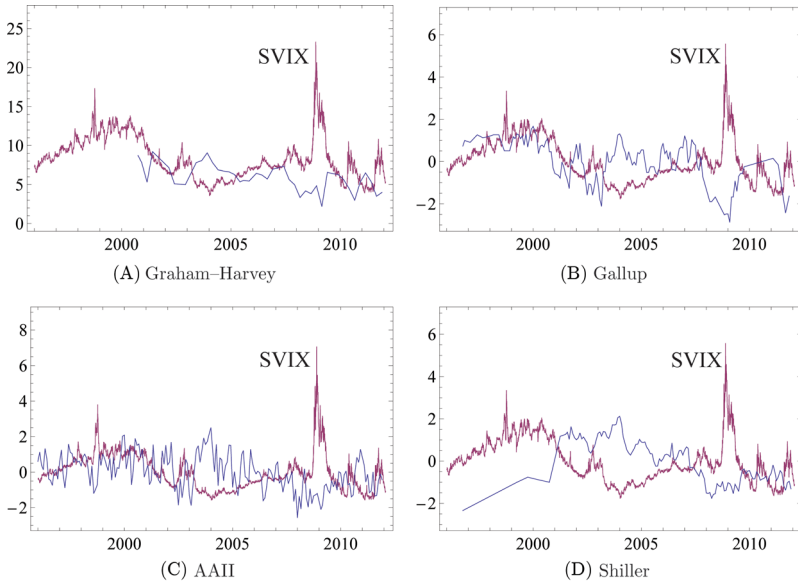
FIGURE VII

Expectations of Returns and SVIX-Implied Expected Returns

The units in Panel A are percentage points. The time series in Panels B, C, and D are normalized to have zero mean and unit variance. The forecasting horizon is one year for Panels A, B, and D, and six months for Panel C.

directly with the expected return implied by SVIX. The other three measures are not in the same units, so Panels B, C, and D of Figure VII show time series standardized to have zero mean and unit variance. The Gallup survey measure is the percentage of investors who are "optimistic" or "very optimistic" about stock returns over the next year, minus the percentage who are "pessimistic" or "very pessimistic." The AAII survey measure is the percentage of surveyed individual investors (members of the AAII) who are "bullish" about stock returns over the next six months, minus the corresponding "bearish" percentage. The Shiller measure reports the percentage of individual investors surveyed who expected the market to go up over the following year.

Each panel also shows the time series of expected returns implied by the SVIX index (calculated by adding the riskless rate to the right-hand side of equation (17)). To be consistent with the phrasing of each survey, I compare the the Gallup, Graham–Harvey, and Shiller surveys to the SVIX-implied equity

TABLE III

CORRELATIONS BETWEEN SURVEY-IMPLIED AND SVIX-IMPLIED EXPECTED RETURNS

	Gallup	Graham–Harvey	AAII	Shiller
Skewness	−0.73	−0.10	0.04	0.06
Excess kurtosis	0.04	−0.28	−0.53	−1.03
Corr(survey, ER)	−0.06	−0.29	−0.20	−0.45
	[.232]	[.030]	[.003]	[.000]
Corr(survey, EER)	−0.53	−0.50	−0.37	−0.46
	[.000]	[.000]	[.000]	[.000]
Corr(Δsurvey, ΔER)	−0.40	−0.21	−0.29	−0.16
	[.000]	[.097]	[.000]	[.035]
Corr(Δsurvey, ΔEER)	−0.44	−0.22	−0.27	−0.16
	[.000]	[.083]	[.000]	[.032]

Notes. Skewness and excess kurtosis of return expectation measures and correlations between return expectations and SVIX-implied expected returns and SVIX-implied expected excess returns (EER) in levels and in differences (denoted by Δ). Numbers in brackets indicate *p*-values on the hypothesis of zero correlation between the two series.

premium (or expected return) at the one-year horizon, and the AAII survey to the six-month SVIX-implied equity premium (or expected return).

It is clear from Figure VII that survey expectations tend to move in the opposite direction from the rational measure of expected returns based on SVIX, as emphasized by Greenwood and Shleifer. As Table III reports, all four survey series are negatively correlated with the SVIX-implied equity premium.[17] This is true whether one measures correlations in levels or in differences, and whether one compares the surveys to the expected return on the market (that is, including the riskless rate, as in the series shown in Figure VII) or to the expected excess return on the market. There is also a contrast in that the skewness and excess kurtosis of the return expectations series are negative or close to zero, whereas they are strongly positive for SVIX, as shown in Table I.[18] Moreover, the lowest points in the Graham–Harvey and Gallup series coincide with the highest point in the SVIX series.

17. I convert the SVIX-implied equity premium into a monthly series by averaging within months, and calculate correlations over all dates that are shared by SVIX and the appropriate survey-based measure.

18. The negative kurtosis of the Gallup and AAII measures may reflect the design of the surveys, each of which provides a fixed scale of possible responses.

Consistent with the thesis of Greenwood and Shleifer, it is implausible, given this evidence, that the surveyed investors have rational expectations.[19] This fact is unsettling for proponents of rational-expectations representative-agent models. (To compound the problem, I show in Section VII.A that none of the leading representative-agent models can match the behavior of VIX and SVIX quantitatively or even qualitatively.) Seen in a certain light, however, this cloud may have a silver lining: the fact that there is a systematic—albeit negative—relationship between (rationally) expected returns and the expectations of surveyed investors points to a pattern that may be amenable to modeling. Barberis et al. (2015) take a first step in this direction by presenting an equilibrium model in which irrational extrapolators interact with rational investors. It is the latter class of investors whose expectations should be thought of as reflected in the SVIX index.

VI. What Is the Probability of a Crash?

The theory presented in Section II was based on a rather minimal assumption: the NCC. I argued in subsequent sections that the NCC may hold with equality, that is, that we may have $\text{cov}_t(M_T R_T, R_T) = 0$. I now strengthen this latter condition further by taking the perspective of an investor with log utility who chooses to invest fully in the market; then the covariance term equals zero for the strong reason that $M_T R_T = 1$. The next result shows how to convert the problem of inferring the subjective expectations of such an investor (written $\widetilde{\mathbb{E}}$, to emphasize that the log investor's viewpoint is taken) into a derivative pricing problem.

RESULT 1. Let X_T be some random variable of interest whose value becomes known at time T, and suppose that we can price a claim to $X_T R_T$ delivered at time T. Then we can compute the

19. On the other hand, Shiller (1987)—reporting the results of investor surveys that were sent out in the immediate aftermath of the crash in October 1987—documents that a substantial fraction of investors expected a market rebound from the crash. Shiller also reports that some investors had more nuanced expectations of market returns: for instance, some thought that the market would perform better over shorter horizons than over long horizons, consistent with the results of Section V.A.

expected value of X_T by pricing an asset:

(22) $\widetilde{\mathbb{E}}_t X_T = $ time t price of a claim to the time T payoff $X_T R_T$.

Proof. A log investor who chooses to invest fully in the market must perceive the market as growth-optimal. The reciprocal of the growth-optimal return is an SDF (Roll 1973), so from the perspective of this log investor, $\frac{1}{R_T}$ is an SDF. The right-hand side of equation (22) therefore equals $\widetilde{\mathbb{E}}_t \left[\frac{1}{R_T} X_T R_T \right]$, which gives the result. $\qquad\square$

I provide two applications in the Appendix: (i) I compare risk-neutral volatility to the real-world volatility perceived by an investor with log utility, and (ii) I show how to modify the proof of Result 1 to compute the expectation of X_T from the perspective of an investor with power utility whose wealth is fully invested in the market, and calculate the equity premium from the perspective of such a investor.

To illustrate how Result 1 can be applied, I now calculate the probability of a market decline from the perspective of the log investor. In this case, the relevant claim can be unambiguously priced because it can be replicated using European index options.[20]

RESULT 2. For simplicity, assume there are no dividend payments between times t and T, so that $R_T = \frac{S_T}{S_t}$. Then the log investor's subjective probability that the return on the market over the period from t to T is less than α is

(23) $\qquad \widetilde{\mathbb{P}}(R_T < \alpha) = \alpha \left[\text{put}'_{t,T}(\alpha S_t) - \frac{\text{put}_{t,T}(\alpha S_t)}{\alpha S_t} \right],$

where $\text{put}'_{t,T}(\cdot)$ is the slope of the put option price curve when plotted as a function of strike.

20. The link between option prices and tail probabilities has been studied by several authors using various different approaches. See, for example, Bates (1991), Backus, Chernov, and Zin (2011), Bollerslev and Todorov (2011), and Barro and Liao (2016).
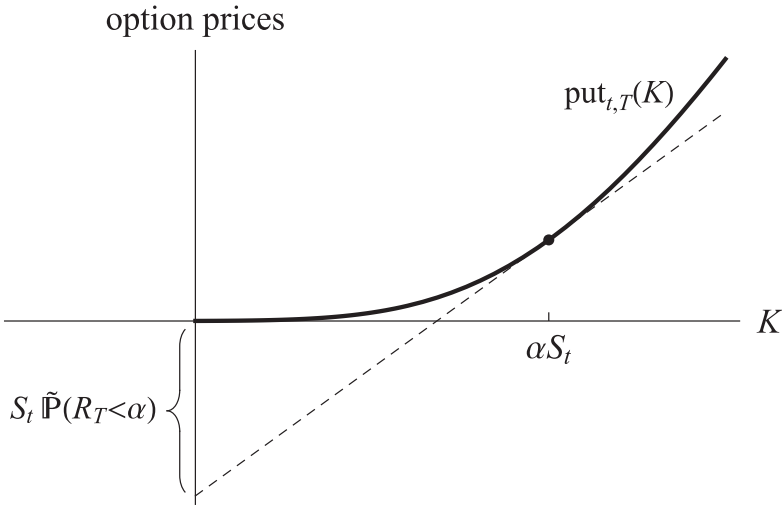
option prices

put$_{t,T}(K)$

αS_t

K

$S_t\,\widetilde{\mathbb{P}}(R_T < \alpha)$

FIGURE VIII

Calculating the Probability of a Crash, $\widetilde{\mathbb{P}}(R_T < \alpha)$, from Put Prices

Proof. Since $\widetilde{\mathbb{P}}(R_T < \alpha) = \widetilde{\mathbb{E}}\left(\mathbf{1}_{\{R_T < \alpha\}}\right)$, we must (by Result 1) price a claim to the payoff $R_T\,\mathbf{1}_{\{R_T < \alpha\}}$. The result follows from the fact that

$$R_T\,\mathbf{1}_{\{R_T < \alpha\}} = \frac{S_T}{S_t}\,\mathbf{1}_{\{S_T < \alpha S_t\}}$$

$$= \alpha\left[\underbrace{\mathbf{1}_{\{S_T < \alpha S_t\}}}_{\text{digital put payoff}} - \frac{1}{\alpha S_t}\underbrace{\max\{0, \alpha S_t - S_T\}}_{\text{put payoff}}\right],$$

since (as is well known and easily checked via a static replication argument) the price of a digital put with strike αS_t—that is, the price of a claim to \$1 paid if and only if $S_T < \alpha S_t$—is put$'_{t,T}(\alpha S_t)$. □

The crash probability index (23) has a geometrical interpretation that is illustrated in Figure VIII: the tangent to put$_{t,T}(K)$ at $K = \alpha S_t$ cuts the y-axis at $-S_t\,\widetilde{\mathbb{P}}(R_T < \alpha)$. Thus the crash probability is high when put prices exhibit significant convexity, as a function of strike, at and below $K = \alpha S_t$.

Panels A–C of Figure IX show the probability of a 20% market crash over the next one, six, and twelve months, smoothed by
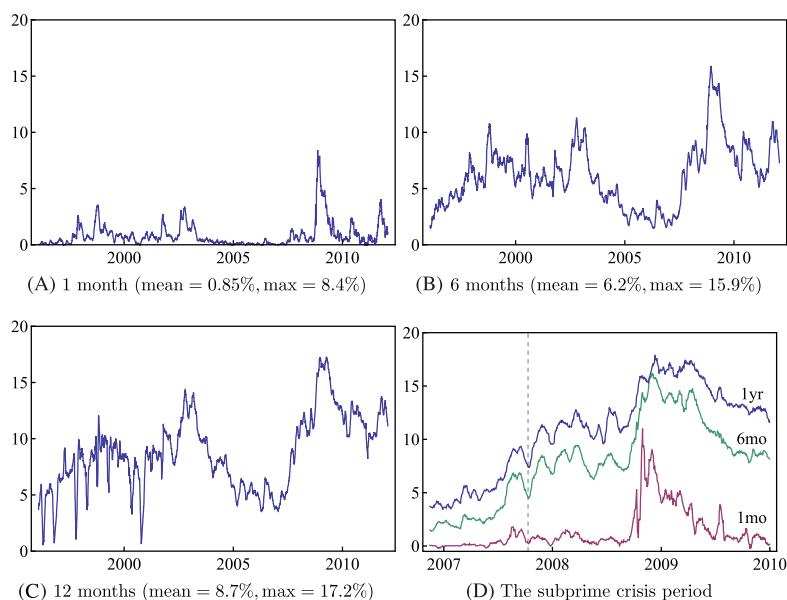
FIGURE IX

The Probability of a 20% Market Crash at Various Horizons

The dashed line in Panel D indicates October 11, 2007, when the S&P 500 index attained an all-time intraday high.

taking a 20-day moving average. The probability of a crash in the next month averages 0.85% over the sample period, and peaks at 8.4%, whereas that of a crash in the next year averages 8.7% and peaks at 17.2%.

Panel D zooms in to show the evolution of the crash probability index at horizons of 1, 6, and 12 months during the subprime crisis.[21] It is notable that the 6-month and 12-month crash probabilities were rising during 2007, before the effects of the subprime crisis started to be felt at the aggregate market level—indeed, at a time when the S&P 500 index itself was rising toward an all-time high on October 11, 2007.

21. The index measures the probability of a market decline of (at least) 20% over a fixed horizon, not the probability of a 20% decline *at any time during* the given horizon; thus, for example, the 12-month crash probability could in principle be lower than the 6-month crash probability.

VII. VIX, SVIX, AND VARIANCE SWAPS

The SVIX index, defined in equation (12), can usefully be compared to the VIX index:

(24)
$$\text{VIX}_{t \to T}^2 \equiv \frac{2R_{f,t}}{T-t} \left\{ \int_0^{F_{t,T}} \frac{1}{K^2} \, \text{put}_{t,T}(K) \, dK + \int_{F_{t,T}}^\infty \frac{1}{K^2} \, \text{call}_{t,T}(K) \, dK \right\}.$$

We saw in equation (13) that the SVIX index measures the risk-neutral volatility of the return on the market. What does VIX measure? Since option prices are equally weighted by strike in the definition of SVIX, but weighted by $\frac{1}{K^2}$ in the definition of VIX, it is clear that VIX places relatively more weight on out-of-the-money puts and less weight on out-of-the-money calls, and hence places more weight on left-tail events.

RESULT 3 (WHAT DOES VIX MEASURE?). If the underlying asset does not pay dividends, so that $R_T = \frac{S_T}{S_t}$, then VIX measures the risk-neutral entropy of the simple return:

(25)
$$\text{VIX}_{t \to T}^2 = \frac{2}{T-t} \, \text{L}_t^* \left(\frac{R_T}{R_{f,t}} \right),$$

where entropy is defined by $\text{L}_t^*(X) \equiv \log \mathbb{E}_t^* X - \mathbb{E}_t^* \log X$.

Proof. As an application of the result of Breeden and Litzenberger (1978), the price of a claim to $\log R_T$ is

$$\frac{1}{R_{f,t}} \mathbb{E}_t^* \log R_T = \frac{\log R_{f,t}}{R_{f,t}} - \int_0^{F_{t,T}} \frac{1}{K^2} \, \text{put}_{t,T}(K) \, dK$$

$$- \int_{F_{t,T}}^\infty \frac{1}{K^2} \, \text{call}_{t,T}(K) \, dK.$$

The result follows by combining this with the fact that $\mathbb{E}_t^* R_T = R_{f,t}$. □

Entropy is a measure of the variability of a positive random variable.[22] Like variance it is nonnegative by Jensen's inequality, and like variance it measures variability by the extent to which a

---

22. 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 (2014).

concave function of an expectation of a random variable exceeds an expectation of a concave function of a random variable.

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. Entropy is more sensitive to the left tail of the return distribution, whereas variance is more sensitive to the right tail, as can be seen by comparing the entropy measure (24), which loads more strongly on out-of-the-money puts, with the variance measure (12), which loads equally on options of all strikes.

The next result shows that VIX and SVIX take a particularly simple form in conditionally log-normal models.

RESULT 4. If the SDF M_T and return R_T are conditionally jointly log-normal, then $\text{SVIX}_{t \to T}^2 = \frac{1}{T-t}(e^{\sigma_t^2(T-t)} - 1)$ and $\text{VIX}_{t \to T}^2 = \sigma_t^2$, where $\sigma_t^2 = \frac{1}{T-t}\text{var}_t \log R_T$. In particular, $\text{SVIX}_{t \to T} > \text{VIX}_{t \to T}$.

Proof. The claims in the first sentence are proved in the Appendix. Since $e^x - 1 > x$ for any real number x, it follows that $\text{SVIX}_{t \to T} > \text{VIX}_{t \to T}$ under log-normality. $\qquad\square$

It would also follow under log-normality that the difference between SVIX and VIX—which the above result shows would be positive—should be negligible for empirically relevant values of σ_t and $T - t$: if for example $\sigma_t = 20\%$ and $T - t = \frac{1}{12}$ (i.e., at a one-month horizon) then we would have $\text{VIX}_{t \to T} = 20\%$ and $\text{SVIX}_{t \to T} = 20.02\%$. Figure X shows (also at a one-month horizon) that these predictions are dramatically violated in the data. The gap between VIX and SVIX is particularly large at times of market stress, but VIX is higher than SVIX on every single day in my sample. This is direct, model-free evidence that the market return and SDF are not conditionally log-normal at the one-month horizon. It is not that non-log-normality only matters at times of crisis; it is a completely pervasive feature of the data. It is also worth emphasizing that this evidence is much stronger than the familiar observation that histograms of log returns are not normal, since that leaves open the possibility that log returns are conditionally normal (with, perhaps, time-varying conditional volatility). Figure X, Panel B excludes that possibility.
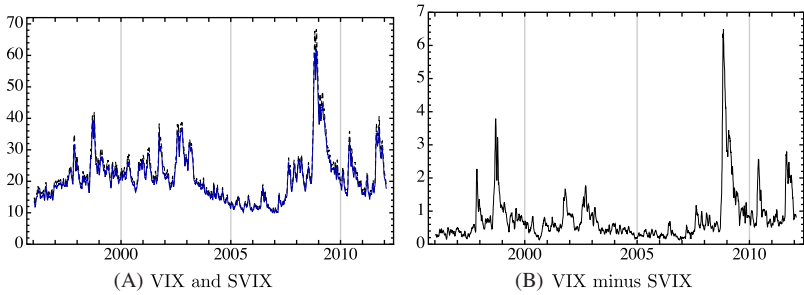
FIGURE X

The Time Series of VIX and SVIX

Panel A: Time series of closing prices of VIX (dotted line) and SVIX (solid line). Panel B: VIX minus SVIX. Both figures show 10-day moving averages.

VII.A. *VIX and SVIX as Diagnostics of Equilibrium Models*

The characterizations of VIX and SVIX in terms of risk-neutral variance and entropy can be read in reverse, as a way to calculate implied VIX and SVIX indexes within equilibrium models: it is easier to calculate risk-neutral entropy and variance than it is to compute option prices and then integrate over strikes.

Can equilibrium models account for the behavior of VIX and SVIX? It might seem that there is room for optimism, given that consumption growth spiked downward in late 2008 as SVIX spiked upward (see Appendix Figure A.3); but as we will now see, leading consumption-based models are unable to match the properties of the two series.

The first three lines of Table IV report various statistics of VIX, SVIX, and VIX minus SVIX: namely, the mean, median, standard deviation, maximum, minimum, skewness, excess kurtosis, and autocorrelation of each series (computed on a monthly basis; full details are provided in the Appendix). The panels report the corresponding quantities calculated within six leading consumption-based models: the Campbell and Cochrane (1999, CC) habit formation model, the long-run risk model in the original stochastic volatility calibration of Bansal and Yaron (2004, BY) and in the more recent calibration of Bansal, Kiku, and Yaron (2012, BKY), the model of Wachter (2013, W) with time-varying disaster arrival rate, and two models that explicitly address the properties of option prices, Bollerslev, Tauchen, and Zhou (2009, BTZ) and Drechsler and Yaron (2011, DY). These numbers are

TABLE IV

SUMMARY STATISTICS OF VIX AND SVIX TIME SERIES IMPLIED BY VARIOUS EQUILIBRIUM MODELS

Data	Mean	Median	Std. dev.	Min	Max	Skewness	Kurtosis	AC(1)
VIX	21.73	20.36	8.54	10.08	67.60	1.91	6.25	0.802
SVIX	20.96	19.77	7.89	9.91	62.31	1.77	5.51	0.803
VIX − SVIX	0.77	0.56	0.75	0.10	6.08	3.60	18.32	0.714
CC								
VIX	18.56	18.72	2.33***	13.46***	22.74***	−0.24***	−0.49**	0.959***
SVIX	18.58	18.75	2.34***	13.47***	22.79***	−0.24**	−0.49**	0.959***
VIX − SVIX	−0.03***	−0.03***	0.01***	−0.05***	−0.01***	−0.09**	−0.28**	0.957***
BY								
VIX	17.20***	17.26*	1.38***	14.08*	19.90***	−0.15***	−0.49***	0.962***
SVIX	17.21**	17.27*	1.38***	14.08*	19.92***	−0.15***	−0.49***	0.962***
VIX − SVIX	−0.01***	−0.01***	0.00***	−0.02***	−0.01***	−0.14***	−0.53***	0.962***
BKY								
VIX	16.64	16.76	2.18***	11.97	20.65***	−0.19***	−0.54***	0.967**
SVIX	16.65	16.77	2.18***	11.98	20.67***	−0.19***	−0.54***	0.967**
VIX − SVIX	−0.01***	−0.01***	0.00***	−0.02***	−0.01***	−0.28***	−0.47***	0.968***
BTZ								
VIX	19.50	19.72	8.63	0.90*	38.02***	−0.08***	−0.50***	0.907*
SVIX	19.53	19.74	8.65	0.90*	38.15**	−0.08***	−0.50***	0.907*
VIX − SVIX	−0.03***	−0.02***	0.03***	−0.13***	0.00***	−1.37***	2.08**	0.911**

TABLE IV
(CONTINUED)

Data	Mean	Median	Std. dev.	Min	Max	Skewness	Kurtosis	AC(1)
DY								
VIX	16.71*	15.27***	4.34*	**13.06*****	37.78**	2.48	7.95	0.825
SVIX	16.62*	15.20***	4.25	**13.06*****	37.33*	2.50	8.05	0.824
VIX − SVIX	**0.09****	**0.06****	**0.09****	**0.00****	**0.45****	1.81**	3.97**	0.838
W								
VIX	38.86	38.79	9.42	19.85	58.25	0.02***	−0.59***	0.966***
SVIX	32.41	32.37	7.69	16.92	48.18	0.02***	−0.59***	0.966***
VIX − SVIX	**6.44****	**6.42****	1.73**	2.93	10.06	**0.04****	**−0.56****	0.965***

Notes. Boldface figures indicate that the value observed in the data lies outside the range generated in the 1,000,000 trials of the given model (so *p*-value < 10⁻⁶); "kurtosis" refers to excess kurtosis, which equals 0 for a normal random variable. *p < .05, **p = .01, ***p = .000 to three decimal places.

generated by simulating 1,000,000 sample paths of VIX and SVIX within each model and computing the average value of the mean, median, and so on across the paths. I also generate an empirical *p*-value for each statistic: this represents the proportion of the 1,000,000 paths that generate values that are as or more extreme as observed in the data.

The results are easily summarized. None of the models comes close to matching the properties of either VIX or SVIX. The difference between the two is particularly problematic: for all six models, the mean (and median) level of VIX minus SVIX observed in the data lies outside the support of the 1,000,000 trials. In the case of the CC, BY, BKY, and BTZ models, which are approximately conditionally log-normal, this failure is a consequence of Result 4. The DY model is not log-normal, but still does not generate a sufficiently large mean gap between VIX and SVIX. The Wachter model, with its extreme disasters, generates too large a mean gap. The models also fail on the other statistics of VIX minus SVIX: its volatility (the Wachter model generates too much, the others not enough), its spikiness (all the models generate too little skewness and kurtosis), and its autocorrelation (higher in the models than in the data). As for VIX and SVIX themselves, only the DY model can match their high skewness and kurtosis and relatively low autocorrelation, and it fails on the other dimensions.

VII.B. *Variance Swaps and Simple Variance Swaps*

The equation underpinning the VIX index (24) is a definition rather than a statement about asset pricing, but the form of the definition originally emerged from the theory of variance swap pricing. This section explores this connection in further detail, and proposes a definition of a tradable contract, the simple variance swap, that is to SVIX as variance swaps are to VIX. As we will see, simple variance swaps are considerably more robust than conventional variance swaps. In particular, they can be hedged even if the underlying asset is subject to jumps. This is an attractive feature, because the variance swap market collapsed during the events of 2008.

In this section we assume that today's date is time 0, and that the goal is to trade (some notion of) variance over the period from

time 0 to time T. A variance swap is an agreement (initiated, say, at time 0) to exchange

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

for some fixed "strike" $\widetilde{V}$ that is defined at time 0 and paid at time T. Here Δ is some small time increment; typically, $\Delta = 1$ day. The market convention is to set $\widetilde{V}$ so that no money needs to change hands at initiation of the trade:

$$
(27) \quad \widetilde{V} = \mathbb{E}_0^* \left[\left(\log \frac{S_\Delta}{S_0}\right)^2 + \left(\log \frac{S_{2\Delta}}{S_\Delta}\right)^2 + \cdots + \left(\log \frac{S_T}{S_{T-\Delta}}\right)^2 \right].
$$

The next result, which is well known, shows how to compute the expectation on the right-hand side of equation (27), under three assumptions that are standard in the variance swap literature but were not required in preceding sections:

A1 the continuously compounded interest rate is constant, at r;
A2 the underlying asset does not pay dividends; and
A3 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 (so that, in particular, there are no jumps).

RESULT 5. Under Assumptions A1–A3, the strike on a variance swap is

$$
(28)
$$
$$
\widetilde{V} = 2e^{rT} \left\{ \int_0^{F_{0,T}} \frac{1}{K^2}\,\mathrm{put}_{0,T}(K)\,dK + \int_{F_{0,T}}^{\infty} \frac{1}{K^2}\,\mathrm{call}_{0,T}(K)\,dK \right\}
$$

in the limit as $\Delta \to 0$; and this quantity has the interpretation

$$
(29) \qquad\qquad \widetilde{V} = \mathbb{E}^* \left[\int_0^T \sigma_t^2\,dt \right].
$$

The variance swap can be hedged by holding
(i) a static position in $\left(\frac{2}{K^2}\right) dK$ puts expiring at time T with strike K, for each $K \leqslant F_{0,t}$,

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

(iii) a dynamic position in $\frac{2\left(\frac{F_{0,t}}{S_t}-1\right)}{F_{0,T}}$ units of the underlying asset at time t, financed by borrowing.

Sketch proof of (28) and (29). In the limit as $\Delta \to 0$, the right-hand side of equation (27) converges to

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

Neuberger (1994) observed that by Itô's lemma and Assumption A3, $d \log S_t = (r - \frac{1}{2}\sigma_t^2)dt + \sigma_t \, dZ_t$ under the risk-neutral measure, so $(d \log S_t)^2 = \sigma_t^2 \, dt$, and

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

$$= 2\,\mathbb{E}^* \left[\int_0^T \frac{1}{S_t} \, dS_t - \int_0^T d \log S_t \right]$$

(30)
$$= 2rT - 2\,\mathbb{E}^* \log \frac{S_T}{S_0}.$$

Carr and Madan (1998) then showed that the approach of Breeden and Litzenberger (1978) can be used to express the price of a claim to the underlying asset's log return in terms of the prices of European call and put options on the underlying asset:

$$P_{\log} \equiv e^{-rT}\,\mathbb{E}^* \log R_T = rTe^{-rT} - \int_0^{F_{0,T}} \frac{1}{K^2}\, \mathrm{put}_{0,T}(K)\,dK$$

$$- \int_{F_{0,T}}^{\infty} \frac{1}{K^2}\, \mathrm{call}_{0,T}(K)\,dK.$$

Substituting back into equation (30), we have the result. $\square$

This result is often referred to as "model-free," since it applies if the underlying asset's price follows any sufficiently well-behaved Itô process. But this is a very strong condition. In reality, the market does not follow an Itô process, so VIX^2 does not

correspond to the fair strike on a variance swap, $\widetilde{V}$;[23] the replicating portfolio provided in Result 5 does not replicate the variance swap payoff; and neither $\widetilde{V}$ nor VIX^2 has the interpretation (29).

Because 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 (26) is *infinite*.

Simple variance swaps do not suffer from this deficiency. A simple variance swap is an agreement to exchange

$$(31) \qquad \left(\frac{S_\Delta - S_0}{F_{0,0}}\right)^2 + \left(\frac{S_{2\Delta} - S_\Delta}{F_{0,\Delta}}\right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{0,T-\Delta}}\right)^2$$

for a prearranged 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.) The choice to put forward prices in the denominators is important: 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 depends on weaker assumptions than were required for the conventional variance swap. Most important, there is no need to assume that the underlying asset follows a diffusion.

23. Aït-Sahalia, Karaman, and Mancini (2015) document a large gap between index variance swap strikes and VIX-type indexes (squared) at all horizons: on the order of 2% in volatility units, compared with an average volatility level around 20%.

B1 the continuously compounded interest rate is constant, at r; and

B2 the underlying asset pays dividends continuously at rate δS_t per unit time.

Given these assumptions, $F_{0,t} = S_0 e^{(r-\delta)t}$. Dividends should be interpreted broadly: if the underlying asset is a foreign currency then δ corresponds to the foreign interest rate. I discuss other ways to deal with dividend payouts in the Appendix.

RESULT 6. (Pricing and hedging a simple variance swap in the $\Delta \to 0$ limit). Under Assumptions B1 and B2, the strike on a simple variance swap is

$$(32) \quad V = \frac{2e^{rT}}{F_{0,T}^2} \left\{ \int_0^{F_{0,T}} \mathrm{put}_{0,T}(K)\, dK + \int_{F_{0,T}}^\infty \mathrm{call}_{0,T}(K)\, dK \right\}.$$

If the asset does not pay dividends, $\delta = 0$, this can be written in terms of the unannualized SVIX formula as

$$V = T \cdot \mathrm{SVIX}_{0 \to T}^2.$$

The payoff on a simple variance swap can be replicated by holding

(i) a static position in $\left(\frac{2}{F_{0,T}^2}\right) dK$ puts expiring at time T with strike K, for each $K \leqslant F_{0,t}$,

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

(iii) a dynamic position in $2e^{-\delta(T-t)} \frac{\left(1 - \frac{S_t}{F_{0,t}}\right)}{F_{0,T}}$ units of the underlying asset at time t, financed by borrowing.

Proof. The derivation of equation (32) divides into two steps. *Step 1*: The absence of arbitrage implies that there are 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, that is,

$$(33) \quad \mathbb{E} \left[M_{(T)} \left\{ \left(\frac{S_\Delta - S_0}{F_{0,0}} \right)^2 + \cdots + \left(\frac{S_T - S_{T-\Delta}}{F_{0,T-\Delta}} \right)^2 - V \right\} \right] = 0.$$

We have

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

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

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

Substituting this into equation (33), we find that

$$(34) \quad V(\Delta) = \sum_{i=1}^{T/\Delta} \frac{e^{ri\Delta}}{F_{0,(i-1)\Delta}^2} \big[\Pi(i\Delta) - (2 - e^{-(r-\delta)\Delta})e^{-\delta\Delta}\Pi((i-1)\Delta)\big].$$

As we have already seen,

$$(35) \quad \Pi(t) = 2\int_0^\infty \mathrm{call}_{0,t}(K)\,dK$$

or, using put-call parity to express $\Pi(t)$ in terms of out-of-the-money options,

$$(36) \quad \Pi(t) = 2\int_0^{F_{0,t}} \mathrm{put}_{0,t}(K)\,dK + 2\int_{F_{0,t}}^\infty \mathrm{call}_{0,t}(K)\,dK + e^{-rt}F_{0,t}^2.$$

Step 2. Observe that equation (34) can be rewritten

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

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 $0 < j < \frac{T}{\Delta}$, the coefficient on $P(j\Delta)$ in this equation is

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

(The definition (31) was originally found by viewing the normalizing constants $F_{0,j\Delta}$, for $j = 0, \ldots \frac{T}{\Delta}$, as arbitrary, and choosing them so that the above equation would hold.) We can therefore rewrite

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

$$(37) \qquad\qquad + \frac{T}{\Delta}\left(e^{(r-\delta)\Delta} - 1\right)^2.$$

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

The trading strategy that replicates the payoff on a simple variance swap is described in detail in the Appendix. $\qquad\square$

The exact expression (34) 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 $\Delta, 2\Delta, \ldots, T$. 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, this concern is misplaced: by choosing forward prices as the normalizing weights in the definition (31), both the pricing formula (34) 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 the definition (31). 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 (though no assumptions have been made about the behavior of the underlying security's price): 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 to offset the effects of increasing delta as calls go in-the-money and puts go increasingly out-of-the-money.

What happens if sampling and trading occurs at discrete intervals $\Delta > 0$, rather than continuously? What if deep-out-of-the-money options cannot be traded? What are the effects of different dividend payout policies? I show in the Appendix that simple variance swaps have good robustness properties in each case.

VIII. CONCLUSION

The starting point of this article is the identity (4), which shows that the expected excess return on any asset equals the risk-neutral variance of the asset's return minus a covariance term. I apply the identity to the return on the market. In this case, risk-neutral variance is directly observable because index options are traded securities, and it turns out that risk-neutral variance is equal to the square of a volatility index, SVIX, that is closely related to the VIX index.

The second term in the decomposition is not directly observable, but it can be controlled: I argue that it is reasonable to assume that the covariance term is negative, and formalize this assumption as the NCC. If the NCC holds, the identity implies that the square of the SVIX index provides a lower bound on the equity premium. I supply various conditions under which the NCC does indeed hold. None of the conditions are necessary, but each is sufficient, so a reader who accepts any one of them—which cover many of the leading equilibrium models of macro-finance, including those of Campbell and Cochrane (1999), Bansal and Yaron (2004), Barro (2006), Wachter (2013), Bansal et al. (2014), and Campbell et al. (2016)—must accept the NCC and the results that flow from it.

I construct the SVIX index using S&P 500 index option data from 1996 to 2012. The index is strikingly volatile: it implies that in late 2008, the equity premium rose above 21% at the one-year horizon and above 55% (annualized) at the one-month horizon. Since the average level of the lower bound over my sample period is approximately 5%, I argue, more aggressively, that the lower bound is approximately tight—that is, risk-neutral variance is not merely a lower bound on the equity premium, it is approximately equal to the equity premium. To assess this possibility, I run predictive regressions and show that the hypothesis of equality is not rejected in the data (though the sample is short, so the standard errors are large). Moreover, when the coefficient on $\text{SVIX}^2_{t \to T}$ is constrained to equal 1—as it is if the bound is tight—the index outperforms the historical mean as a return predictor; Goyal and Welch (2008) showed that this is not true of a range of predictor variables that have been proposed in the literature. These empirical results suggest that the SVIX index can be used as a direct proxy for the equity premium. I do so, and draw various conclusions.

First, the equity premium is far more volatile than implied by the valuation-ratio predictors of Campbell and Thompson (2008). The distinction between the two views is sharpest on days such as Black Monday, in October 1987, when the S&P 500 and Dow Jones indexes experienced very severe declines, with daily returns roughly twice as negative as the next-worst day in history. On the Campbell–Thompson view of the world, the equity premium rose on the order of 2 or 3 percentage points during this episode. In sharp contrast, option prices are known to have exploded on Black Monday, which I argue implies also that the equity premium exploded.

Second, the equity premium is strongly right-skewed: the median equity premium is on the order of 4%, but there are occasional opportunities for unconstrained investors to earn a much higher equity premium.

Third, during such episodes, a disproportionate fraction of the equity premium is concentrated in the form of extremely high expected returns over the very short run.

Fourth, while I have argued that the SVIX index proxies for the rationally expected return on the market, it is, like other predictor variables that have been proposed in the literature, negatively correlated with survey measures of investor return expectations (which forecast returns with the wrong sign).

Fifth, the equity premium exhibits significant fluctuations at monthly, weekly, daily, and still higher frequencies. The macro-finance literature, which seeks to rationalize market gyrations at the business cycle frequency, typically has not acknowledged or attempted to address such movements. The basic point can be made by focusing on the behavior of the VIX and SVIX indexes (thereby making the argument independent of the NCC). I compute the two indexes within a range of leading equilibrium models and find that none is able to generate sample paths of VIX and SVIX that even roughly resemble the data: the models generate sample paths that are more persistent and less spiky than the paths of the indexes observed in reality. This casts further doubt on the plausibility of such models, quite apart from the fact that they cannot be squared with survey evidence.

Since models featuring a representative agent who consumes aggregate consumption are inconsistent with the data in each of these last two respects, I prefer to take the perspective of a rational, unconstrained investor whose wealth is invested in the market—or in assets that are well proxied by the market—and whose risk aversion is at least 1, so that the NCC holds. (As shown in Section III, more general assumptions are possible: if the investor's portfolio is less than fully invested in the market, the NCC holds subject to a stronger restriction on risk aversion.) This approach allows for the possibility that this rational, unconstrained investor coexists with irrational and/or constrained investors. But since there are many ways to be irrational and many ways to be constrained, the equity premium perceived by a rational unconstrained investor is the natural benchmark (even though such an investor, whose beliefs are embedded in asset prices, is not necessarily representative of the full population: as an extreme example, a tightly constrained investor may have essentially arbitrary beliefs without these beliefs being detectable in asset prices).

The article concludes with extensions on two rather different dimensions. In the first, I consider the problem of inferring the subjective expectations of an unconstrained log investor who chooses to hold the market. This represents a special case in which the NCC holds with equality, and the inference problem then reduces to an exercise in derivative pricing. As an application, I compute the probability of a market crash from option prices. As one might have expected, the resulting crash probability index tends to be high at times when the equity premium is high; more interestingly, the one-year crash probability started to rise in late

2007, at a time when the S&P 500 index was itself at historic highs, and well before the unfolding subprime crisis led to sharp declines in the stock market.

The second extension completes a square. The definition of the VIX index is motivated by the theory of variance swap pricing. I show how to define a contract—the simple variance swap—that is to SVIX as variance swaps are to VIX. If such a contract were traded in liquid markets, it would directly reveal the level of SVIX in real time, obviating the need to observe option prices at all strikes. Moreover, simple variance swaps have the advantage that they can be priced and hedged even in the presence of jumps in the underlying asset price. This is not true of conventional variance swaps, and as a result the variance swap market collapsed during the market turmoil of 2008.

APPENDIX

This section contains proofs that the examples in Section III satisfy the NCC.

Example 1. Write $M_T = e^{-r_{f,t} + \sigma_{M,t} Z_{M,T} - \frac{\sigma_{M,t}^2}{2}}$ and $R_T = e^{\mu_{R,t} + \sigma_{R,t} Z_{R,T} - \frac{\sigma_{R,t}^2}{2}}$, where $Z_{M,T}$ and $Z_{R,T}$ are (potentially correlated) standard normal random variables. The requirement that $\mathbb{E}_t\, M_T R_T = 1$ implies that $\mu_{R,t} - r_{f,t} + \mathrm{cov}_t(\log M_T, \log R_T) = 0$. This fact, together with some straightforward algebra, implies that $\mathbb{E}_t\, M_T R_T^2 \leqslant \mathbb{E}_t\, R_T$ if and only if $\lambda_t \geqslant \sigma_{R,t}$, where λ_t is the conditional Sharpe ratio $\frac{\mu_{R,t} - r_{f,t}}{\sigma_{R,t}}$.

Example 2. By assumption, there is an investor with wealth W_t and utility function $u(\cdot)$ who chooses, at time t, from the available menu of assets with returns $R_T^{(i)}$, $i = 1, 2, \ldots$. In other words, he chooses portfolio weights $\{w_i\}$ to solve the problem

$$(38) \qquad \max_{\{w_i\}} \mathbb{E}_t\, u\left[W_t \left(\sum_i w_i R_T^{(i)} \right) \right] \quad \text{subject to} \quad \sum_i w_i = 1.$$

The first-order condition for (say) w_j is that

$$\mathbb{E}_t\left[W_t u'\left(W_t \sum_i w_i R_T^{(i)} \right) R_T^{(j)} \right] = \lambda_t,$$

where $\lambda_t > 0$ is the Lagrange multiplier associated with the constraint in (38). Since the investor chooses to hold the market, we have $\sum_i w_i R_T^{(i)} = R_T$. Thus,

$$\mathbb{E}_t \left[\underbrace{\frac{W_t}{\lambda_t} u'(W_t R_T)}_{M_T} R_T^{(j)} \right] = 1$$

for any return $R_T^{(j)}$. It follows that the SDF is proportional (with a constant of proportionality that is known at time t) to $u'(W_t R_T)$.

To show that the NCC holds, we must show that $\text{cov}_t(u'(W_t R_T)R_T, R_T) \leqslant 0$. This holds because $u'(W_t R_T)R_T$ is decreasing in R_T: its derivative is $u'(W_t R_T) + W_t R_T u''(W_t R_T) = -u'(W_t R_T)[\gamma(W_t R_T) - 1]$, which is negative because risk aversion $\gamma(x) \equiv -\frac{xu''(x)}{u'(x)}$ is at least 1.

If the investor has log utility, then $\gamma(x) \equiv 1$, so the inequality holds with equality. But it is not *necessary* for the investor to have log utility for the inequality to hold with equality: all we require is that $M_T R_T$ is uncorrelated with R_T. That is, we merely need that $M_T = I_T/R_T$ where I_T and R_T are uncorrelated (and $\mathbb{E}_t I_T = 1$ since $\mathbb{E}_t M_T R_T$ must equal 1). Log utility is the special case in which $I_T \equiv 1$.

Examples 3a and 3b. For reasons given in the text, Example 3a is a special case of Example 3b, which we now prove. We must check that $\text{cov}_t(M_T R_T, R_T) \leqslant 0$, or equivalently that

$$(39) \qquad \text{cov}_t(-R_T V_W(W_T, z_{1,T}, \ldots, z_{N,T}), R_T) \geqslant 0.$$

So we must prove that the covariance of two functions of $R_T, R_T^{(i)}, z_{1,T}, \ldots, z_{N,T}$ is positive. The two functions are

$$f(R_T, R_T^{(i)}, z_{1,T}, \ldots, z_{N,T}) = -R_T V_W\big(\alpha_t(W_t - C_t)R_T$$
$$+ (1 - \alpha_t)(W_t - C_t)R_T^{(i)}, z_{1,T}, \ldots, z_{N,T}\big)$$

and

$$g(R_T, R_T^{(i)}, z_{1,T}, \ldots, z_{N,T}) = R_T.$$

(Since the covariance is conditional on time t information, α_t and $(W_t - C_t)$ can be treated as known constants.) By the defining

property of associated random variables, (39) holds if f and g are each weakly increasing functions of their arguments. This is obviously true for g, so we must check that the first derivatives of f are all nonnegative.

Differentiating f with respect to R_T, we need $-V_W(W_T, z_{1,T}, \ldots, z_{N,T}) - \alpha_t(W_t - C_t)R_T V_{WW}(W_T, z_{1,T}, \ldots, z_{N,T}) \geqslant 0$, or equivalently

$$-\frac{W_T V_{WW}(W_T, z_{1,T}, \ldots, z_{N,T})}{V_W(W_T, z_{1,T}, \ldots, z_{N,T})} \geqslant \frac{W_T}{W_{M,T}},$$

where W_T and $W_{M,T}$ are as given in the main text; this is the constraint on risk aversion. Differentiating with respect to $R_T^{(i)}$, we need $-R_T(1 - \alpha_t)(W_t - C_t)V_{WW}(W_T, z_{1,T}, \ldots, z_{N,T}) \geqslant 0$, which follows because $V_{WW} < 0$. Differentiating with respect to $z_{j,T}$, we need $-R_T V_{Wj}(W_T, z_{1,T}, \ldots, z_{N,T}) \geqslant 0$, which follows because V_{Wj} (the cross derivative of the value function with respect to wealth and the jth state variable) is weakly negative given the choice of sign on the state variables.

Examples 4a and 4b. With Epstein–Zin preferences, the SDF is proportional (up to quantities known at time t) to $\frac{\left(\frac{W_T}{C_T}\right)^{\frac{(\gamma-1)}{(1-\psi)}}}{R_T^{-\gamma}}$, so the desired inequality, $\text{cov}_t(M_T R_T, R_T) \leqslant 0$, is equivalent to

$$\text{cov}_t\left[-\left(\frac{W_T}{C_T}\right)^{\frac{(\gamma-1)}{(1-\psi)}} R_T^{1-\gamma}, R_T\right] \geqslant 0.$$

If $\gamma = 1$, as in Example 4b, then this holds with equality.

If $\frac{W_T}{C_T}$ and R_T are associated, as assumed in Example 4a, then we need to check that the first derivatives of $f(x, y) = -x^{\frac{(\gamma-1)}{(1-\psi)}}y^{1-\gamma}$ are nonnegative. But this holds if $\gamma \geqslant 1$ and $\psi \geqslant 1$, as claimed.

A. Construction of the Lower Bound

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. 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 data points. I compute mid-market option prices by averaging the highest closing bid and lowest closing ask, and using the resulting prices to compute the lower bound by discretizing the right-hand side of inequality (14).

On any given day, I compute the lower bound at 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 seven days; this is the same procedure the CBOE follows. I then calculate the bound 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 seven days, requiring me to use the two expiry dates further out; again, this is the procedure followed by the CBOE.

B. The Effect of Discrete Strikes

The integrals that appear throughout the article are idealizations: in practice we only observe options at some finite set of strikes. Write $\Omega_{t,T}(K)$ for the price of an out-of-the-money option with strike K, that is,

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

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 $F_{t,T}$;[24] and define $\frac{(K_{i+1} - K_{i-1})}{2}$. Then the idealized integral $\int_0^\infty \Omega_{t,T}(K)\,dK$ is replaced, in practice, by the observable sum $\sum_{i=1}^N \Omega_{t,T}(K_i)\,\Delta K_i$. (This is the CBOE's procedure in calculating VIX, and I follow it in this article.) Appendix Figure A.2, Panel A illustrates.

RESULT 7 (The effect of discretization by strike). Discretizing by strike will tend to lead to an underestimate of the idealized

24. I assume for simplicity 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.
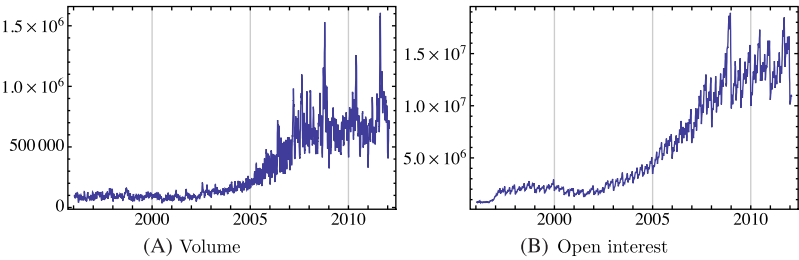
FIGURE A.1

Volume and Open Interest in S&P 500 Index Options
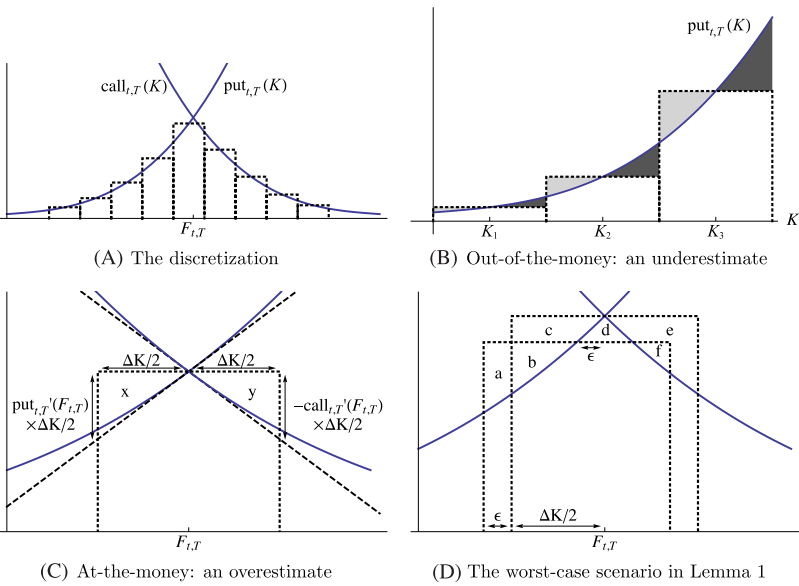
The figures show 10-day moving averages.

FIGURE A.2

The Effect of Discretization

Different panels use different scales.

lower bound, in that

$$\underbrace{\frac{2}{(T-t)R_{f,t}S_t^2}\sum_{i=1}^{N}\Omega_{t,T}(K_i)\,\Delta K_i}_{\text{discretization}} \leqslant \underbrace{\frac{2}{(T-t)R_{f,t}S_t^2}\int_0^{\infty}\Omega_{t,T}(K)\,dK}_{\text{idealized lower bound}}$$

$$+\underbrace{\frac{(\Delta K_j)^2}{4(T-t)\cdot R_{f,t}^2\cdot S_t^2}}_{\text{very small}}.$$

Proof. Nonobservability of deep-out-of-the-money options obviously leads to an underestimate of the lower bound.

Consider, first, the out-of-the-money puts with strikes $K_1, \ldots, K_{j-1}$. The situation is illustrated in Appendix Figure A.2, Panel B: by convexity of $\text{put}_{t,T}(K)$, the light gray areas that are included (when they should be excluded) are smaller than the dark gray 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 $F_{t,T}$, as in Appendix Figure A.2, Panel C. 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 $\text{put}_{t,T}(K)$ and $\text{call}_{t,T}(K)$ at $K = F_{t,T}$. The areas of the resulting triangles provide the desired upper bound, by convexity of $\text{put}_{t,T}(K)$ and $\text{call}_{t,T}(K)$:

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

But by put-call parity, $\text{put}'_{t,T}(K) - \text{call}'_{t,T}(K) = \frac{1}{R_{f,t}}$. Thus, the overestimate due to the at-the-money option is at most $\frac{1}{2}\left(\frac{\Delta K}{2}\right)^2 \frac{1}{R_{f,t}}$. 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 scales the integral by $\frac{2}{((T-t)R_{f,t}S_t^2)}$, the result follows. □

The maximal overestimate provided by this result is very 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 $F_{t,T} = 1000$ then at a monthly horizon, $T - t = \frac{1}{12}$, 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 I. Since the nonobservability of deep-out-of-the-money options causes underestimation, there is 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 7. The goal is to consider the largest possible overestimate that the option whose strike is nearest to the forward price, $F_{t,T}$, can contribute. Appendix Figure A.2, Panel D illustrates. The dotted rectangle in the figure is the contribution if the strike happens to be equal to $F_{t,T}$; I call this Case 1. The dashed rectangle is the contribution if the strike equals $F_{t,T} - \varepsilon$, for some $\varepsilon > 0$ (the case $\varepsilon < 0$ is essentially identical); I 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 area (b) + area (c) + area (e) + area (f) $\geqslant$ area (a) + area (b) + area (f) − area (d) in Figure A.2 Panel D, i.e., if area(c) + area(d) + area(e) $\geqslant$ area(a). But by convexity of $\mathrm{put}_{t,T}(K)$, area(b) + area(c) $\geqslant$ area(a) + area(b), which gives the result. An almost identical argument applies if $\varepsilon < 0$. □

C. Supplementary Tables and Figures

Appendix Figure A.3, Panel A shows that the VXO index of one-month at-the-money implied volatility on the S&P 100 index exploded on October 19, 1987. Panel B plots log real consumption growth (quarterly, seasonally adjusted personal consumption expenditures, from the Bureau of Economic Analysis) and the one-year SVIX-implied equity premium on the same axes, with each time series scaled to have zero mean and unit variance. Panel C plots the rolling mean historical equity premium, computed using

(A) VXO

(B) Consumption growth and SVIX

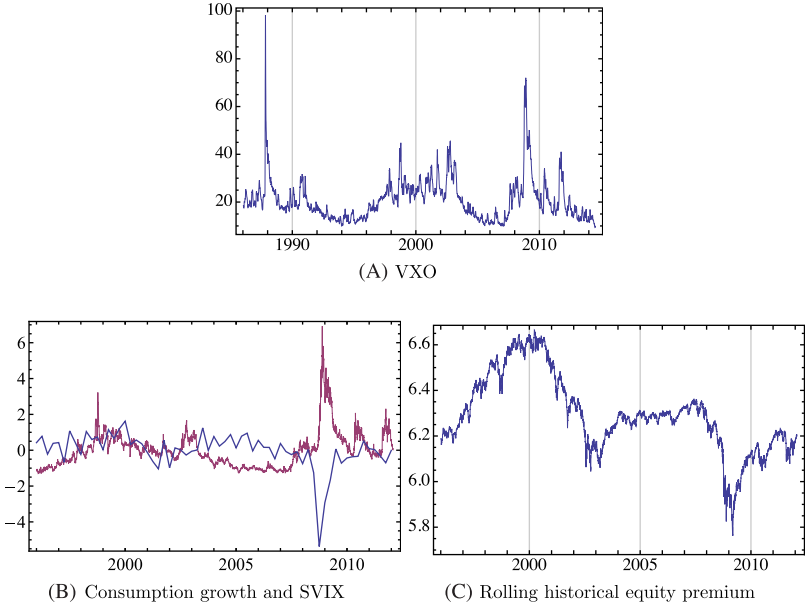(C) Rolling historical equity premium

FIGURE A.3

Supplementary Figures

TABLE A.1

COEFFICIENT ESTIMATES FOR THE REGRESSION (16), EXCLUDING THE CRISIS PERIOD
AUGUST 1, 2008–JULY 31, 2009 FROM THE SAMPLE

Horizon	$\widehat{\alpha}$	Std. err.	$\widehat{\beta}$	Std. err.	R^2 (%)
1 mo	−0.095	0.061	3.705	1.258	3.36
2 mo	−0.081	0.062	3.279	1.181	4.83
3 mo	−0.076	0.067	3.147	1.258	5.98
6 mo	−0.043	0.072	2.319	1.276	4.94
1 yr	0.045	0.088	0.473	1.731	0.27

the data series used by Campbell and Thompson (2008) based on
S&P 500 total returns from February 1871, with the data prior to
January 1927 obtained from Robert Shiller's website.

Table A.1 reproduces the results in Table II, but excludes the
period August 1, 2008–July 31, 2009. Table A.2 reports results for
regressions

$$(40) \quad R_T - R_{f,t} = \alpha + \beta_1 \times R_{f,t} \cdot \text{SVIX}^2_{t \to T} + \beta_2 \times \text{VRP}_{t \to T} + \varepsilon_T$$

TABLE A.2

COEFFICIENT ESTIMATES FOR THE REGRESSION (40)

Horizon	$\widehat{\alpha}$	Std. err.	$\widehat{\beta_1}$	Std. err.	$\widehat{\beta_2}$	Std. err.	R^2 (%)
1 mo	−0.086	0.063	2.048	1.273	3.908	1.053	4.96
2 mo	−0.113	0.061	2.634	1.007	3.884	0.761	8.54
3 mo	−0.086	0.071	2.273	1.407	2.749	0.346	6.79
6 mo	−0.051	0.076	1.992	1.132	−0.525	1.259	6.56
1 yr	−0.073	0.078	2.278	0.909	−0.694	0.680	10.34

(A) 1-month   (B) 1-year
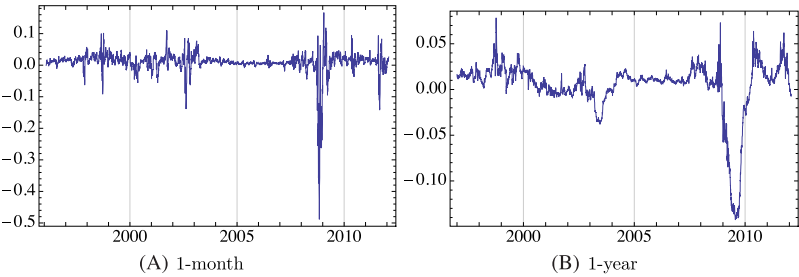
FIGURE A.4

The Variance Risk Premium, Calculated as $R_{f,t} \cdot \text{SVIX}^2_{t \to T} - \text{SVAR}_{t \to T}$

of realized returns onto risk-neutral variance and a measure of the variance risk premium, $\text{VRP}_{t \to T} \equiv R_{f,t} \cdot \text{SVIX}^2_{t \to T} - \text{SVAR}_{t \to T}$. Realized daily return variance, $\text{SVAR}_{t \to T}$, is computed at time t by looking backward over the same horizon length, $T - t$, as the corresponding forward-looking realized return (so, for example, I use one-month backward-looking realized variances to predict one-month forward-looking realized returns). If realized variance is a good proxy for forward-looking real-world variance, this is a measure of the variance risk premium.

Consistent with the findings of Bollerslev, Tauchen, and Zhou (2009) and Drechsler and Yaron (2011), the coefficient on VRP_t is positive and strongly significant at predictive horizons out to three months. This predictive success reflects the fact that implied and realized volatility, $\text{SVIX}_{t \to T}$ and $\text{SVAR}_{t \to T}$, rose sharply as the S&P 500 dropped in late 2008; implied volatility then fell relatively quickly, while $\text{SVAR}_{t \to T}$ declined more sluggishly. $\text{VRP}_{t \to T}$ therefore turned dramatically negative in late 2008, as shown in Appendix Figure A.4. Since the market then continued

TABLE A.3

COEFFICIENT ESTIMATES FOR THE REGRESSION (40), EXCLUDING THE CRISIS PERIOD AUGUST 1, 2008, TO JULY 31, 2009

Horizon	$\widehat{\alpha}$	Std. err.	$\widehat{\beta_1}$	Std. err.	$\widehat{\beta_2}$	Std. err.	$R^2(\%)$
1 mo	−0.103	0.061	3.333	1.292	1.548	1.125	3.61
2 mo	−0.097	0.063	3.137	1.353	1.532	1.801	6.04
3 mo	−0.083	0.068	2.902	1.451	1.133	1.855	6.34
6 mo	0.016	0.071	0.797	1.560	0.360	2.095	0.74
1 yr	0.008	0.061	0.331	2.274	1.761	3.760	3.10

to fall, this sluggish response of $\text{VRP}_{t \to T}$ helps fit the data. At the six-month and one-year horizons, however, $\text{VRP}_{t \to T}$ responds too sluggishly—it remains strongly negative even as the market starts to rally in March 2009—so there is a sign-flip, with negative estimates of the coefficient on $\text{VRP}_{t \to T}$ at the six-month and one-year horizons. The empirical facts are therefore hard to interpret: the sign flip raises the concern that the apparent success of $\text{VRP}_{t \to T}$ as a predictor variable may be an artefact of this particular sample period. Table A.3 therefore repeats the regression (40), but excludes the period from August 1, 2008, to July 31, 2009. Once this crisis period is excluded, $\text{VRP}_{t \to T}$ does not enter significantly at any horizon.

From a theoretical point of view, it is hard to rationalize a negative equity premium forecast within any equilibrium model. It is also implausible that the correctly measured variance risk premium should ever be negative. More specifically, Bollerslev, Tauchen, and Zhou (2009) show that within their own preferred equilibrium model, the variance risk premium would always be positive.

D. VIX, SVIX, and Equilibrium Models

Proof of Result 4. I write $\tau = T - t$ to make the notation easier to handle. Let $R_T = e^{\mu_{R,t}\tau + \sigma_t\sqrt{\tau}Z_R - \frac{\sigma_t^2\tau}{2}}$ and $M_T = e^{-r_{f,t}\tau + \sigma_{M,t}\sqrt{\tau}Z_M - \frac{\sigma_{M,t}^2\tau}{2}}$, where Z_R and Z_M are normal random variables with mean 0, variance 1, and correlation ρ_t, and $r_{f,t} = \log R_{f,t}$. Since $\mathbb{E}_t M_T R_T = 1$, we must have $\mu_{R,t} - r_{f,t} = -\rho_t \sigma_{M,t} \sigma_t$.

From equation (13), $\text{SVIX}^2_{t\to T} = \frac{e^{-2r_{f,t}\tau}}{\tau}\left[\mathbb{E}^*_t(R^2_T) - (\mathbb{E}^*_t R_T)^2\right] = \frac{1}{\tau}\left(e^{-r_{f,t}\tau}\,\mathbb{E}_t\,M_T R^2_T - 1\right)$. Now, using the fact that $\mu_{R,t} - r_{f,t} = -\rho_t\sigma_{M,t}\sigma_t$, we have

$$\mathbb{E}_t\,M_T R^2_T = \mathbb{E}\,e^{-r_{f,t}\tau+\sigma_{M,t}\sqrt{\tau}Z_M - \frac{1}{2}\sigma^2_{M,t}\tau + 2\mu_{R,t}\tau + 2\sigma_t\sqrt{\tau}Z_R - \sigma^2_t\tau}$$

$$= e^{r_{f,t}\tau + \sigma^2_t\tau}.$$

Thus, $\text{SVIX}^2_{t\to T} = \frac{1}{\tau}(e^{\sigma^2_t\tau} - 1)$ as required.

The calculation for VIX is slightly more complicated. Using equation (25), $\text{VIX}^2_{t\to T} = \frac{2}{\tau}\,\text{L}^*_t R_T = \frac{2}{\tau}\left[\log\mathbb{E}^*_t R_T - \mathbb{E}^*_t \log R_T\right] = \frac{2}{\tau}\left[r_{f,t}\tau - e^{r_{f,t}\tau}\,\mathbb{E}_t\,M_T\log R_T\right]$. Now,

$$\mathbb{E}_t\left[M_T\log R_T\right] = \mathbb{E}_t\left[\left(\mu_{R,t}\tau + \sigma_t\sqrt{\tau}Z_R - \frac{1}{2}\sigma^2_t\tau\right)\right.$$

$$\left.\times\; e^{-r_{f,t}\tau + \sigma_{M,t}\sqrt{\tau}Z_M - \frac{\sigma^2_{M,t}\tau}{2}}\right]$$

$$= \left(\mu_{R,t} - \frac{1}{2}\sigma^2_t\right)\tau\cdot e^{-r_{f,t}\tau}$$

$$+ \sigma_t\sqrt{\tau}e^{-r_{f,t}\tau - \frac{\sigma^2_{M,t}\tau}{2}}\,\mathbb{E}_t\left[Z_R e^{\sigma_{M,t}\sqrt{\tau}Z_M}\right].$$

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

$$\mathbb{E}_t\left[Z_R e^{\sigma_{M,t}\sqrt{\tau}Z_M}\right] = \mathbb{E}_t\left[(\rho_t Z_M + \sqrt{1-\rho^2_t}Z)e^{\sigma_{M,t}\sqrt{\tau}Z_M}\right]$$

$$= \rho_t\,\mathbb{E}_t\left[Z_M e^{\sigma_{M,t}\sqrt{\tau}Z_M}\right].$$

By Stein's lemma,

$$\mathbb{E}_t\left[Z_M e^{\sigma_{M,t}\sqrt{\tau}Z_M}\right] = \mathbb{E}_t\left[\sigma_{M,t}\sqrt{\tau}e^{\sigma_{M,t}\sqrt{\tau}Z_M}\right]$$

$$= \sigma_{M,t}\sqrt{\tau}e^{\frac{\sigma^2_{M,t}\tau}{2}}.$$

It follows (using the fact that $\mu_{R,t} - r_{f,t} = -\rho_t\sigma_{M,t}\sigma_t$) that $\mathbb{E}_t\,M_T\log R_T = (\mu_{R,t} - \frac{1}{2}\sigma^2_t + \rho_t\sigma_{M,t}\sigma_t)\tau e^{-r_{f,t}\tau} = (r_{f,t} - \frac{1}{2}\sigma^2_t)\tau e^{-r_{f,t}\tau}$. So $\text{VIX}^2_{t\to T} = \sigma^2_t$, as required. $\square$

The top panel of Table IV reports a variety of summary statistics for VIX, SVIX, and VIX minus SVIX in the data, at the one-month horizon. (For comparison with the models, which are simulated at monthly frequency, I generate a monthly series from the daily series of VIX and SVIX by taking the 1st, 22nd, 43rd, 64th, ..., elements and compute the mean, median, etc. Then I repeat using the 2nd, 23rd, ... elements; the 3rd, 24th, ...; and so on, up to the 21st, 42nd, ... Finally, I average each statistic over the 21 choices of initial element.) The panels below report corresponding statistics computed within the equilibrium models of Campbell and Cochrane (1999, CC), Bansal and Yaron (2004, BY), Bansal, Kiku, and Yaron (2012, BKY), Wachter (2013, W), Bollerslev, Tauchen, and Zhou (2009, BTZ), and Drechsler and Yaron (2011, DY).

Within each model, I simulate 1,000,000 16-year-long sample paths of VIX, SVIX, and VIX minus SVIX. Each sample path is generated by initializing state variables at their long-run averages, then computing a 32-year sample realization. I discard the first 16-year "burn-in" period and use the second 16 years (for comparability with the 16 years of data). I compute VIX and SVIX at the one-month horizon for all models apart from Wachter's continuous-time model, for which I compute an instantaneous measure, that is, I report the limiting case in which the time horizon $T - t$ approaches 0, rather than one month.[25] Along each

---

25. I do so for tractability, because this quantity can be computed in closed form. Since the Wachter model generates too little skewness and kurtosis in VIX and SVIX, and too much persistence, it is likely that using the one-month measure would make the results even worse. It is also possible to solve for VIX and SVIX in closed form within Barro's model; in this case, VIX and SVIX are constant and their term structures are flat, so there is no distinction between the instantaneous and one-month measures. Following Martin (2013) by defining $\kappa(\theta) \equiv \log \mathbb{E}_t \left[\left(\frac{C_{t+1}}{C_t} \right)^\theta \right]$, it can be shown that within Barro's (2006) model—or indeed within any consumption-based model with an Epstein–Zin representative agent and i.i.d. consumption growth—we have

$$\text{VIX}^2 = 2\big[\kappa(1-\gamma) - \kappa(-\gamma) - \kappa'(-\gamma)\big]$$

$$\log\left(1 + \text{SVIX}^2\right) = \kappa(2-\gamma) - 2\kappa(1-\gamma) + \kappa(-\gamma),$$

where γ is relative risk aversion. Using the calibration of Barro (2006), one finds that VIX = 23.8% and SVIX = 18.4%, and that the difference between the two is 5.4%, well above the value observed in the data. (These calculations assume that there is no default on index options, and model equity as an unlevered claim to consumption. Allowing for default would move the numbers in the right

sample path, I compute the mean, standard deviation, median, minimum, maximum, skewness, excess kurtosis, and monthly autocorrelation for VIX, SVIX, and VIX minus SVIX. The numbers reported in Table IV are the averages, across 1,000,000 sample paths in each model, of each of these quantities. Asterisks in Table IV indicate statistics for which a given model struggles to match the data. The models find it so difficult to match the data that I use fewer asterisks than is conventional to indicate significance levels: one asterisk denotes a *p*-value of .05 (fewer than 5% of the 1,000,000 trials gave statistics as extreme as are observed in the data), two asterisks a *p*-value of .01, three asterisks a *p*-value of .000 to three decimal places. Boldface font indicates that the observed statistic in the data lies completely outside the support of the 1,000,000 model-generated statistics (i.e., an empirical *p*-value of 0). Obviously a successful model should not have any boldface statistics; unfortunately, there are multiple such examples for all six models.

E. Simple Variance Swaps

The proof of Result 6 implicitly supplies the dynamic trading strategy that replicates the payoff on a simple variance swap. Tables A.4 and A.5 describe the strategy in detail. Each row of Table A.4 indicates a sequence of dollar cash flows 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 (35); this portfolio has simple return $\frac{S_\Delta^2}{\Pi(\Delta)}$ 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

---

direction—the gap between VIX and SVIX would decline—since VIX loads more heavily on the deep out-of-the-money puts that would be most vulnerable to default. Allowing for leverage would move the numbers in the wrong direction, expanding the gap between VIX and SVIX.)

TABLE A.4

REPLICATING THE SIMPLE VARIANCE SWAP.

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

Notes. In the left column, B indicates dollar positions in the bond, U indicates dollar positions in the underlying with dividends continuously reinvested, and $j\Delta$, for $j = 1, 2, \ldots$, $\frac{T}{\Delta}$, 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 $\Pi_{j\Delta}$

TABLE A.5

REPLICATING THE SIMPLE VARIANCE SWAP: THE GENERIC POSITION IN OPTIONS OF INTERMEDIATE MATURITY, TOGETHER WITH THE ASSOCIATED TRADES REQUIRED AFTER EXPIRY.

Asset	0	$j\Delta$	$(j+1)\Delta$	T
$j\Delta$	$\dfrac{-(e^{(r-\delta)\Delta}-1)^2\Pi_{j\Delta}}{e^{r(T-j\Delta)}F_{0,j\Delta}^2}$	$\dfrac{(e^{(r-\delta)\Delta}-1)^2 S_{j\Delta}^2}{e^{r(T-j\Delta)}F_{0,j\Delta}^2}$		
B		$e^{-r(T-j\Delta)}\left[\dfrac{-S_{j\Delta}^2}{F_{0,(j-1)\Delta}^2} - \dfrac{S_{j\Delta}^2}{F_{0,j\Delta}^2}\right]$		$\dfrac{S_{j\Delta}^2}{F_{0,(j-1)\Delta}^2} + \dfrac{S_{j\Delta}^2}{F_{0,j\Delta}^2}$
U		$\dfrac{2S_{j\Delta}^2 e^{-\delta\Delta}}{e^{r(T-(j+1)\Delta)}F_{0,j\Delta}^2}$	$\dfrac{-2S_{j\Delta}S_{(j+1)\Delta}}{e^{r(T-(j+1)\Delta)}F_{0,j\Delta}^2}$	
B			$\dfrac{2S_{j\Delta}S_{(j+1)\Delta}}{e^{r(T-(j+1)\Delta)}F_{0,j\Delta}^2}$	$\dfrac{-2S_{j\Delta}S_{(j+1)\Delta}}{F_{0,j\Delta}^2}$

Notes. 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 $j\Delta$

underlying asset, and subsequently from 2Δ to T in the bond. The replicating portfolio requires similar positions in options expiring at times 2Δ, 3Δ, ..., $T - 2\Delta$. These are omitted from Table A.4, but the general such position is indicated in Table A.5, 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 0. The last column of Table A.4 adds up to the desired payoff (31) minus the strike V. The first column must therefore add up to the cost of entering the simple variance swap. Equating this cost to 0, we find the value of V provided in equation (34).

The replicating strategy simplifies nicely in the $\Delta \to 0$ limit. The dollar investment in each of the option portfolios expiring at times Δ, 2Δ, ..., $T - \Delta$ goes to 0 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 A.5, this calls for a short position in the underlying asset of $\frac{2e^{-r(T-(j+1)\Delta)}S_{j\Delta}^2 e^{-\delta\Delta}}{F_{0,j\Delta}^2}$ in dollar terms at time $j\Delta$, that is, a short position of $\frac{2e^{-r(T-(j+1)\Delta)}S_{j\Delta}e^{-\delta\Delta}}{F_{0,j\Delta}^2}$ units of the underlying. In the limit as $\Delta \to 0$, holding $j\Delta = t$ constant, this equates to a short position of $\frac{2e^{-r(T-t)}S_t}{F_{0,t}^2}$ units of the underlying asset at time t.

The static position in options expiring at time T, shown in the penultimate line of Table VIII, 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 equation (35). 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 $\frac{2e^{-\delta(T-t)}}{F_{0,t}}$ 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

$$\frac{2e^{-\delta(T-t)}}{F_{0,T}} - \frac{2e^{-r(T-t)}S_t}{F_{0,t}^2} = \frac{2e^{-\delta(T-t)}(1-\frac{S_t}{F_{0,t}})}{F_{0,T}}$$ units of the asset and long out-of-the-money-forward calls and puts, all financed by borrowing.

London School of Economics

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

References

Aït-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, "The Term Structure of Variance Swaps and Risk Premia," Working Paper, 2015.

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

Ang, Andrew, and Geert Bekaert, "Stock Return Predictability: Is it There?," *Review of Financial Studies*, 20 (2007), 651–707.

Backus, David, Mikhail Chernov, and Ian Martin, "Disasters Implied by Equity Index Options," *Journal of Finance*, 66 (2011), 1969–2012.

Backus, David, Mikhail Chernov, and Stanley Zin, "Sources of Entropy in Representative Agent Models," *Journal of Finance*, 69 (2014), 51–99.

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

Bansal, Ravi, Dana Kiku, and Amir Yaron, "An Empirical Evaluation of the Long-run Risks Model for Asset Prices," *Critical Finance Review*, 1 (2012), 183–221.

Bansal, Ravi, Dana Kiku, Ivan Shaliastovich, and Amir Yaron, "Volatility, the Macroeconomy, and Asset Prices," *Journal of Finance*, 69 (2014), 2471–2511.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, "X-CAPM: An Extrapolative Capital Asset Pricing Model," *Journal of Financial Economics*, 115 (2015), 1–24.

Barro, Robert J., and Gordon Liao, "Options-pricing Formula with Disaster Risk," Technical report, *NBER*, 2016.

Barro, Robert J., "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121 (2006), 823–866.

Bates, David S., "The Crash of '87: Was it Expected? The Evidence from Options Markets," *Journal of Finance* 46 (1991), 1009–1044.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey, "Managerial Mis-calibration," *Quarterly Journal of Economics*, 128 (2013), 1547–1584.

Black, Fischer, and Myron Scholes, "The Pricing of Options and Corporate Liabil-ities," *Journal of Political Economy*, 81 (1973), 637–654.

Bollerslev, Tim, and Viktor Todorov, "Tails, Fears, and Risk Premia," *Journal of Finance*, 66 (2011), 2165–2211.

Bollerslev, Tim, George Tauchen, and Hao Zhou, "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22 (2009), 4463–4492.

Breeden, Douglas T., and Robert H. Litzenberger, "Prices of State-contingent Claims Implicit in Option Prices," *Journal of Business*, 51 (1978), 621–651.

Campbell, John Y., Stefano Giglio, Christopher Polk, and Robert Turley, "An In-tertemporal CAPM With Stochastic Volatility," Working Paper, 2016.

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

Campbell, John Y., and Robert J. Shiller, "The Dividend-Price Ratio and Expecta-tions of Future Dividends and Discount Factors," *Review of Financial Studies*, 1 (1988), 195–228.

Campbell, John Y., and Samuel B. Thompson, "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?," *Review of Finan-cial Studies*, 21 (2008), 1509–1531.

Campbell, John Y., and Luis M. Viceira, "Consumption and Portfolio Decisions When Expected Returns Are Time-varying," *Quarterly Journal of Economics*, 114 (1999), 433–495.

Carr, Peter, and Dilip Madan, "Towards a Theory of Volatility Trading," in *Volatil-ity: New Estimation Techniques for Pricing Derivatives*, Robert Jarrow ed. (London: Risk Books, 1998).

Cochrane, John H., *Asset Pricing* (Princeton, NJ: Princeton University Press, 2005).

——, "The Dog that Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, 21 (2008), 1533–1575.

——, "Presidential Address: Discount Rates," *Journal of Finance*, 66 (2011), 1047–1108.

Drechsler, Itamar, and Amir Yaron, "What's Vol Got to Do with It," *Review of Financial Studies*, 24 (2011), 1–45.

Epstein, Larry G., and Stanley E. Zin, "Substitution, Risk Aversion, and the Tem-poral Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57 (1989), 937–969.

——, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99 (1991), 263–286.

Esary, J. D., F. Proschan, and D. W. Walkup, "Association of Random Variables, with Applications," *Annals of Mathematical Statistics*, 38 (1967), 1466–1474.

Fama, Eugene F., and Kenneth R. French, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22 (1988), 3–25.

——, "The Equity Premium," *Journal of Finance*, 57 (2002), 637–659.

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

Greenwood, Robin, and Andrei Shleifer, "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27 (2014), 714–746.

Hansen, Lars P., and Robert J. Hodrick, "Forward Exchange Rates as Optimal Pre-dictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88 (1980), 829–853.

Hansen, Lars P., and Ravi Jagannathan, "Implications of Security Market Data For Models of Dynamic Economies," *Journal of Political Economy*, 99 (1991), 225–262.

Keim, Donald B., and Robert F. Stambaugh, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics*, 17 (1986), 357–390.

Kelly, Bryan, and Seth Pruitt, "Market Expectations in the Cross-Section of Present Values," *Journal of Finance*, 68 (2013), 1721–1756.

Lettau, Martin, and Sydney Ludvigson, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56 (2001), 815–849.

Malmendier, Ulrike, and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?," *Quarterly Journal of Economics*, 126 (2011), 373–416.

Martin, Ian, "Consumption-based Asset Pricing with Higher Cumulants," *Review of Economic Studies*, 80 (2013), 745–773.

Merton, Robert C., "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41 (1973), 867–887.

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

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

Pitt, Loren D., "Positively Correlated Normal Variables Are Associated," *Annals of Probability*, 10 (1982), 496–499.

Roll, Richard, "A Critique of the Asset Pricing Theory's Tests I: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4 (1977), 129–176.

——, "Evidence on the 'Growth-Optimum' Model," *Journal of Finance*, 28 (1973), 551–566.

Ross, Stephen A., *Neoclassical Finance* (Princeton, NJ: Princeton University Press, 2005).

Shiller, Robert J., "Investor Behavior in the October 1987 Stock Market Crash: Survey Evidence," Technical report, *NBER*. 1987.

——, *Irrational Exuberance* (Princeton, NJ: Princeton University Press, 2000).

Stambaugh, Robert F., "Predictive Regressions," *Journal of Financial Economics*, 54 (1999), 375–421.

Wachter, Jessica A., "Can Time-varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68 (2013), 987–1035.

Weil, Philippe, "Nonexpected Utility in Macroeconomics," *Quarterly Journal of Economics*, 105 (1990), 29–42.