Information in Derivatives Markets: Forecasting Prices with Prices

Ian W. R. Martin

London School of Economics, London WC2A 2AE, UK, i.w.martin@lse.ac.uk.

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–27

https://doi.org/10.1146/((please add article doi))

Keywords

market forecasting, return predictability, derivatives, options

Abstract

I survey work that uses information in derivative and other asset prices to forecast movements in financial markets.

Contents

In this paper, I survey a recent literature that uses information in derivative and other asset prices to forecast movements in financial markets. This literature aims to provide answers to questions such as:

1. What is the expected return on the market over the next six months?
2. What is the expected return on Apple stock over the next year?
3. By how much is the euro expected to appreciate or depreciate against the dollar over the next two years?
4. What is the probability that Apple stock drops by 30% over the next quarter?
5. How do long-run expectations currently compare with short-run expectations?
6. What is the expected future path of interest rates?
7. What is the expected inflation rate?
8. What is the expected growth rate of aggregate dividends?
9. How autocorrelated are market returns expected to be?
10. ...and so on.

There is, of course, no single answer to any of these questions. Almost any survey, formal or informal, will elicit a range of responses to each one; different econometricians will come up with different "objective" measures of conditional expectations. Some people—even perhaps some economists—will give answers that seem obviously false to other people. At best, we might hope to come up with answers to these questions that could be broadly accepted as reasonable.

I will emphasize various distinctive, and interrelated, features of the literature I survey.

First, minimal assumptions are made about the underlying price processes. It is common in the asset pricing literature to assume that asset prices and returns are lognormally distributed, or that they follow diffusion processes. These assumptions lead to tractable models, but they are not plausible in reality.

Second, measures of volatility come in at least three different flavors: true, risk-neutral, and historical. In lognormal models all three are essentially the same. In general, they are all different. I will emphasize measures of implied volatility based on option prices. In a lognormal world, option prices are uninteresting, determined passively from the underlying asset (as in the groundbreaking paper of Black and Scholes (1973)). In general, however, the dynamic replication of options or other derivatives is impossible, so that they must be viewed as genuinely distinct assets priced in equilibrium rather than by the pure theory of no arbitrage—and their prices convey genuinely distinct information.

Third, the theory is set up with measurement in mind. The classical theory of financial economics relates risk premia to conditional covariances (with the return on the market in the capital asset pricing model (CAPM) or with risk factors in the arbitrage pricing theory). These quantities are not observable in practice, so it is conventional to proxy for them with historical measures of realized covariance. But, as Martin and Wagner (2019) put it, "when markets are turbulent, historical betas may not accurately reflect the idealized forward-looking betas called for by the CAPM, or by factor models more generally; and if the goal is to forecast returns over, say, a one-year horizon, one cannot respond to this critique by taking refuge in the last five minutes of high-frequency data." The papers surveyed here relate risk premia to risk-neutral variances, covariances and other risk-neutral quantities that are directly observable from forward-looking asset prices.

Section 1 introduces the formula of Merton (1980) that connects the market's risk premium to its variance, and discusses some extensions. In Section 2, I show how the beliefs of a representative investor with log utility can be inferred from asset prices. I derive connections between the market's risk premium and its risk-neutral variance, and between arbitrary assets' risk premia and their risk-neutral covariances with the market, and show how these quantities can be calculated from appropriate derivative prices. In Section 3, I derive an identity which holds without any assumptions on the form of the SDF, and use it to generalize the approach beyond the log investor. Section 4 concludes.

1. MERTON'S FORMULA

Merton (1969, 1971) considered the problem of how an individual with power utility should invest in an iid world with a fixed riskless rate r_f and a risky asset whose price, S_t, follows a geometric Brownian motion,

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dZ. \qquad 1.$$

The optimal share of wealth allocated to the risky asset, α, is

$$\alpha = \frac{\mu - r_f}{\gamma\sigma^2}, \qquad 2.$$

where $\mu - r_f$ is the expected excess return[1] on the risky asset and γ is the coefficient of relative risk aversion.

If all individuals have the same level of risk aversion, and if the risky asset is interpreted as the market portfolio, equilibrium requires that $\alpha = 1$, so that equation 2 provides a relationship between the expected excess return on the market and the volatility of the market:

$$\mu - r_f = \gamma\sigma^2. \qquad 3.$$

The appeal of this relationship, which was derived and analyzed by Merton (1980), is that the expected return on the market is hard to estimate directly from time series data, whereas market volatility is easy to calculate in the geometric Brownian motion setting.

More generally, if individuals with different levels of risk aversion γ_i face the price process described by equation 1, they will choose different risky shares $\alpha_i = (\mu - r_f)/(\gamma_i\sigma^2)$.

[1]The analog of $\mu - r_f$ in a discrete time model is $\log \mathbb{E}_t \frac{R_{t+1}}{R_{f,t+1}}$, where R_{t+1} is the one-period gross risky return and $R_{f,t+1}$ is the one-period gross riskless rate.

In this case, the equilibrium requirement is that the wealth-weighted-average risky portfolio allocation should equal one: writing w_i for individual i's wealth share, we must have $\sum_i w_i \alpha_i = 1$. This implies that

$$\mu - r_f = \overline{\gamma}\sigma^2, \tag{4.}$$

where $\overline{\gamma} = \left(\sum_i \frac{w_i}{\gamma_i}\right)^{-1}$ is wealth-weighted harmonic mean risk aversion.

Two aspects of this aggregate risk aversion measure deserve emphasis. First, wealthy individuals receive more weight in the calculation of $\overline{\gamma}$. Observers wishing to understand the behavior of financial markets should devote particular attention to the risk preferences of the rich.

Second, the harmonic mean is particularly sensitive to the presence of individuals with low risk aversion. If there are two equally wealthy individuals with risk aversion 1 and 1,000,000, respectively, then arithmetic mean risk aversion is slightly more than 500,000 and geometric mean risk aversion is 1,000; and yet harmonic mean risk aversion, $\overline{\gamma}$, is slightly less than two! People with low risk aversion have a disproportionate influence on financial markets because they trade aggressively. Carried to the extreme, the presence of even one unconstrained and truly risk-neutral agent ($\gamma_i = 0$) drives aggregate risk aversion to zero. Conversely, someone with infinite risk aversion will not participate in risky financial markets, and so will have almost no impact on the pricing of risky assets.

In this GBM setting, volatility, σ, can be calculated either by computing realized quadratic variation directly from the price process using high-frequency data over any finite time interval or by observing option prices. Indeed we could use the price of an option with *any* strike and *any* time to maturity: as the Black and Scholes (1973) model would hold, all options would have the same implied volatility which—like the expected return on the market under the model assumed in equation 1—would be constant over time.

In practice, the empirical literature has tended to use rolling measures of realized volatility to proxy for forward-looking conditional volatility, with results that are typically only weakly supportive of the basic equation 3.[2]

1.1. Jumps and lumpy information

Casual observation quickly reveals, however, that asset prices do not follow geometric Brownian motions. Volatility moves around over time; and prices jump discontinuously, sometimes at unexpected times (a terrorist attack occurs, a major bank fails, a war breaks out) and sometimes at predictable times (an economic number is released, an election takes place).[3] Consequently, implied volatility, inferred from option prices on some fixed underlying asset, varies with strike and time to maturity, and over time, and is itself subject to jumps.[4]

The idea that information can arrive in lumps is a fundamental challenge to the Brownian motion view of the world. As Karatzas and Shreve (1998) write, at the start of their

[2]See, for example, Merton (1980), French et al. (1987), Baillie and DeGennaro (1990), Campbell and Hentschel (1992), and Harvey (2001). In the other direction, Guo and Whitelaw (2006), Ludvigson and Ng (2007) and Pastor et al. (2008) argue for a positive relationship between risk and return. In a different style, Bekaert et al. (2009) present evidence supportive of a positive relationship between risk and return in the context of a consumption-based asset pricing model.

[3]See, for example, Aït-Sahalia (2002), Chernov et al. (2003), Barndorff-Nielsen and Shephard (2004), Todorov (2009), and Aït-Sahalia and Jacod (2012).

[4]See, for example, Bates (1991), Carr and Wu (2003), and Broadie et al. (2007).

textbook on financial markets driven by Brownian motions: "Our assumption that asset prices have no jumps is a significant one. It is tantamount to the assertion that there are no 'surprises' in the market".

But manifestly there are 'surprises' in the market. Several recent papers formally document the impact of major macroeconomic announcements on asset prices, consistent with the lumpy information view of the world.[5]

The analysis above can be adapted to accommodate the case of a jump that takes place at a known point in time, if we imagine that, at time t, the market will be multiplied by a lognormal random variable $J = e^{\left(\mu-\frac{1}{2}\sigma^2\right)+\sigma Z}$, where Z is standard Normal. To induce a representative investor with power utility to continue to invest fully in the market over a short interval around time t, we must have $\mu - r_f = \gamma\sigma^2$ as before. (This follows from a direct calculation, or simply by noting that the units of time in equations 1–3 are indeterminate so that we can choose them as we wish, provided that μ, σ and r_f are interpreted appropriately.) Suppose, for example, that at the time of an announcement known to be occurring in the next ten seconds, the market price will experience (i.e., be multiplied by) a lognormal jump with standard deviation $\sigma = 2\%$. In this case, the equilibrium expected excess return would be $\gamma \times (2\%)^2 = 4\gamma$ basis points *over the next ten seconds.*[6] This order of magnitude is consistent with results reported by Savor and Wilson (2013), who find that the average excess return on major macroeconomic news announcement days from 1958 to 2009 is 11.4 basis points.

To illustrate what happens when news arrives unexpectedly, suppose that the asset price experiences jumps at times determined by a Poisson process with arrival rate ω:

$$\frac{dS_t}{S_t} = \mu\, dt - L\,(dN - \omega\, dt)\,. \tag{5}$$

To keep things simple, suppose that when news arrives, the jump is of fixed size L, where $L > 0$, so that a jump represents bad news, and $L < 1$, so that the asset's price always remains positive. This is a highly stylized example, but it is an important counterpoint to the GBM specification of equation 1.

The pure jump model has two free parameters, ω and L, to compare with the single parameter σ in the Brownian case. To put them on the same footing, we choose ω and L so that volatility is the same in each case, i.e., we set $\omega = \sigma^2/L^2$. We can imagine fixing volatility, σ, and then choosing the parameter L freely. Large values of L correspond to rare extreme disasters, whereas values of L close to zero correspond to frequent small jumps. We can think of information arriving occasionally in large lumps if L is large; or arriving frequently in small pieces when L is close to zero. The optimal share of wealth invested in the risky asset is then[7]

$$\alpha_{\text{jump}} = \frac{1}{L}\left[1 - \left(\frac{\sigma^2}{(\mu - r_f)L + \sigma^2}\right)^{1/\gamma}\right]. \tag{6}$$

[5] See, for example, Savor and Wilson (2013), Savor and Wilson (2014), Lucca and Moench (2015), Ai and Bansal (2018), Cieslak et al. (2019), and Hillenbrand (2024). Backus et al. (2011) argue that the jumps whose influence is evident in option prices should be thought of as frequently occurring small jumps rather than large rare disasters of the type emphasized by Barro (2006).

[6] Moreover, as the 10-second riskless return is approximately zero, the 4γ bp expected excess return in the example is almost exactly equal to the expected return.

[7] See the NBER working paper by Campbell and Martin (2021) for a derivation.

Imposing the requirement that $\alpha_{\text{jump}} = 1$ in equilibrium, equation 6 implies that

$$\mu - r_f = \frac{(1-L)^{-\gamma} - 1}{L}\sigma^2. \qquad 7.$$

This is the analog of the Merton formula given in equation 3. Expanded as a power series in L, equation 7 becomes

$$\mu - r_f = \gamma\sigma^2 + \frac{\gamma(\gamma+1)L}{2!}\sigma^2 + \frac{\gamma(\gamma+1)(\gamma+2)L^2}{3!}\sigma^2 + \frac{\gamma(\gamma+1)(\gamma+2)(\gamma+3)L^3}{4!}\sigma^2 + \cdots. \qquad 8.$$

In the limit as L tends to zero—with small pieces of information arriving very frequently—this simplifies to the Merton formula shown in equation 3. More generally, though, the equity premium depends not only on risk aversion and variance but also on higher moments of the asset return (determined here by the jump size and arrival rate, as captured by L). At first sight this is discouraging, as estimating the stochastic properties of jumps that may only occur infrequently is econometrically challenging.

2. THE LOG INVESTOR

Given a traded payoff X_{t+1}, the time t risk-neutral expectation of X_{t+1} is defined to be

$$\mathbb{E}_t^* X_{t+1} = R_{f,t+1}\,\mathbb{E}_t\left(M_{t+1}X_{t+1}\right). \qquad 9.$$

Although risk-neutral expectations are often[8] used as a rough guide to true expectations, equation 9 shows that the two types of expectations are only identical in the special, and counterfactual, case in which pricing is *genuinely risk-neutral*—that is, only if

$$M_{t+1} = 1/R_{f,t+1}. \qquad 10.$$

This is, undeniably, a crude assumption. And yet the appeal of risk-neutral quantities reflects the fact that—as they can be inferred from asset prices alone, without the need for infrequently updated macroeconomic or accounting data—they are observable in real time. A second advantage is that there are no free parameters to be estimated or calibrated: to give one example, breakeven (that is, risk-neutral) inflation is an unambiguous quantity on which market participants can agree whatever their personal views on the macroeconomy.

In this section, I discuss an approach that has echoes of the Merton formula, but which (like the risk-neutral approach) makes no assumptions about the stochastic processes fol-

[8]Notably, practitioners do not use risk-neutral expected returns as approximations to true expected returns. The reason is that the risk-neutral expected return on any asset equals the riskless rate. As risk premia are large relative to riskless rates for most asset classes of interest, risk-neutral expected returns have not been useful measures of "market-implied expected returns." (If we lived in a world with high and widely fluctuating interest rates, risk-neutral expected returns might come to seem a more sensible measure.) Nor can they differentiate cross-sectionally, with one exception: if exchange rates are involved then there are multiple different riskless rates in play, one for each currency. And, indeed, in this context, the risk-neutral approach does have an interesting role: the risk-neutral expected appreciation of one currency relative to another is determined by the two currencies' interest rates. That is, the risk-neutral forecast equals the uncovered interest parity (UIP) forecast, a quantity which is often viewed as a benchmark in the international finance literature.

lowed by asset prices, other than that they are arbitrage-free.[9] Specifically, I adopt the perspective of an unconstrained, rational, marginal investor with log utility over next-period wealth. This individual may coexist with other (rational or irrational) individuals with different preferences and/or different beliefs, but we assume that he or she chooses to hold the market.[10] As we will see, it is possible to use derivative prices to infer the perceptions such an investor must have about (for example) expected returns on the market, on currencies, and on other assets.[11]

I study this simple case throughout this section for several reasons. First, it represents a useful benchmark with no free parameters that exhibits the main ideas in a particularly simple way. Second, it has the pedagogical advantage that the resulting expressions have echoes of familiar relationships that arise in traditional models: for instance an asset's risk premium is proportional to its (risk-neutral) covariance with the market. Third, the discussion around equation 4 motivates the choice of a utility function with relatively low risk aversion. Fourth, it helps to emphasize that a *single* model makes coherent predictions across a range of asset classes: we will have a comprehensive view of "the world according to the log investor." Fifth, utility should properly be defined over real quantities: thus we should think of the log investor as maximizing expected log real return. But expected log real returns decompose nicely—$\mathbb{E}_t \log \frac{R_{t+1}}{\pi_{t+1}} = \mathbb{E}_t \log R_{t+1} - \mathbb{E}_t \log \pi_{t+1}$—so we can simply think of the investor as maximizing expected log nominal returns, $\mathbb{E}_t \log R_{t+1}$, and work in nominal terms throughout.

The resulting theory can be generalized in several ways: for example, via an identity that generalizes the key equation 13 below, or by deriving bounds that relax the exact equalities of this section. We can also make different assumptions on the rational investor whose perspective is taken. For example, for some applications it is easy to allow the investor to have an arbitrary utility function; alternatively, we might continue to think from the perspective of a log investor, but allow for the possibility that he or she chooses to hold an asset other than the market. I discuss these and other issues in Section 3.

Today is time t, and we suppose that there is an investor operating in the market who has log utility over next-period wealth. If this investor is rational, he or she solves the problem

$$\max_{w_1,\dots,w_N} \mathbb{E}_t \left[\log \sum_{i=1}^N w_i R_{i,t+1} \right] \quad \text{s.t.} \sum_{i=1}^N w_i = 1. \qquad 11.$$

The associated first-order conditions are that

$$\mathbb{E}_t \left[\frac{R_{j,t+1}}{\sum_{i=1}^N w_i R_{i,t+1}} \right] = 1 \quad \text{for all } j \in \{1,\dots,N\}. \qquad 12.$$

[9]Santa-Clara and Yan (2010) take an approach that is similar in spirit, but impose considerably more structure, estimating an equilibrium model featuring stochastic volatility and stochastic jump intensity. The model yields a forecasting relationship that expresses the equity premium in terms of the model's diffusive volatility and jump intensity, each of which is inferred from option prices.

[10]Martin and Papadimitriou (2022) present an equilibrium model with heterogeneous beliefs in which, at every point in time, there is a representative log investor who holds the market; but the degree of optimism of the representative investor shifts depending on market conditions.

[11]For a related approach, see Bliss and Panigirtzoglou (2004), who use option prices from a somewhat different angle, using option prices to make inferences about the representative agent's relative risk aversion.

We assume that the investor is marginal in all assets, so that the first-order conditions have an interior solution. The collection of equations 12 then shows that the reciprocal of the investor's chosen portfolio return is an SDF.

Suppose that in equilibrium the investor holds the S&P 500, which we think of as a proxy for the idealized "market portfolio" that comes out of theory, i.e. the market-cap-weighted portfolio of all assets in positive supply.[12] Writing R_{t+1} for the gross return on the S&P 500, we then have $\sum_{i=1}^N w_i R_{i,t+1} = R_{t+1}$, so that $M_{t+1} = 1/R_{t+1}$. I neglect the effect of dividends, writing $R_{t+1} = S_{t+1}/S_t$. In some of the cases considered below this could be replaced by an assumption that dividends are known one period ahead so that for example $\operatorname{var}_t^* R_{t+1} = \operatorname{var}_t^* \frac{S_{t+1}+D_{t+1}}{S_t} = \operatorname{var}_t^* \frac{S_{t+1}}{S_t}$.

Suppose now that we want to infer the log investor's expectations about some variable X_{t+1}. As

$$\mathbb{E}_t X_{t+1} = \mathbb{E}_t \left(\frac{X_{t+1} R_{t+1}}{R_{t+1}} \right) = \underbrace{\frac{1}{R_{f,t+1}} \mathbb{E}_t^* \left(X_{t+1} R_{t+1} \right)}_{\text{price of a claim to } X_{t+1} R_{t+1}} \qquad 13.$$

this converts the belief inference problem to a derivative pricing problem: if we can observe, or can calculate, the price of a claim to $X_{t+1}R_{t+1}$, then we can infer the log investor's expectations about X_{t+1}.

2.1. The market

2.1.1. The expected return on the market. Setting $X_{t+1} = R_{t+1}$ in equation 13,

$$\mathbb{E}_t R_{t+1} = \frac{1}{R_{f,t+1}} \mathbb{E}_t^* \left(R_{t+1}^2 \right). \qquad 14.$$

As the risk-neutral expectation of any asset's return equals the riskless rate, we find a relationship between the log investor's expected excess return and the *risk-neutral* variance of the market:

$$\mathbb{E}_t R_{t+1} - R_{f,t+1} = \frac{1}{R_{f,t+1}} \operatorname{var}_t^* R_{t+1}. \qquad 15.$$

This equation, which was first derived in Martin (2011), is reminiscent of Merton's formula (3) specialized to the case $\gamma = 1$; but, unlike Merton's formula, it does not require assumptions on the underlying price process. It also applies if, say, the market experiences stochastic volatility, or jumps as in the example of Section 1.1. In the presence of jumps, risk-neutral variance can be very different from true variance, as equation 7 implicitly shows.

Risk-neutral variance has the great advantage that it is directly observable from option prices. More generally, Breeden and Litzenberger (1978) showed that it is possible to calculate risk-neutral expectations of the form $\mathbb{E}_t^* g(S_{t+1})$ for any random variable S_{t+1}—usually an asset price—on which European-style options are traded. For, assuming $g(\cdot)$ is a suitably well-behaved function, we have the relationship

$$g(S_{t+1}) = g(F) + g'(F)(S_{t+1} - F) + \\ + \int_0^F g''(K) \max\{0, K - S_{t+1}\}\, dK + \int_F^\infty g''(K) \max\{0, S_{t+1} - K\}\, dK. \qquad 16.$$

[12] By market-clearing, the wealth-weighted average investor must hold the market portfolio. By making the assumption that the investor holds the S&P 500 index explicit, we are acknowledging the force of the Roll (1977) critique.

Table 1: Forecasting the market. Daily data.

$\frac{R_{t+1}}{R_{f,t+1}} - 1 = \alpha + \beta \, \mathrm{SVIX}_t^2 + \varepsilon_{t+1}$						
	1996.01–2022.12			2012.02–2022.12		
horizon	α	β	R^2 (%)	α	β	R^2 (%)
1mo	0.014 [0.041]	1.569 [1.024]	1.302	-0.015 [0.035]	4.280 [0.822]	7.809
3mo	0.015 [0.051]	1.439 [1.269]	2.201	-0.008 [0.041]	3.620 [0.997]	12.334
6mo	-0.025 [0.037]	2.418 [0.806]	6.840	0.011 [0.053]	3.007 [1.633]	11.454
12mo	0.003 [0.043]	1.858 [0.824]	4.727	0.008 [0.070]	3.188 [2.368]	11.316

SVIX is constructed using S&P 500 index option price data from OptionMetrics. Total returns are calculated from CRSP daily returns. Observations are daily. Newey–West standard errors (with 21, 65, 130, and 260 lags at horizons of 1, 3, 6 and 12 months, respectively) are reported in square brackets.

There is no economics here: this is simply an equation reminiscent of a Taylor expansion, but the second-order terms are weighted integrals over option-like payoffs.

We can allow F to be an arbitrary constant, but it will now be convenient to set it equal to the time $t+1$ forward price of the asset, F_t, which is chosen to make the value of the forward trade equal to zero at initiation: thus $\mathbb{E}_t^* \left(S_{t+1} - F_t\right) = 0$. Using this fact, taking conditional risk-neutral expectations and discounting by the riskless rate, the price of a claim to $g(S_{t+1})$ is

$$\frac{1}{R_{f,t+1}} \mathbb{E}_t^* \, g(S_{t+1}) = \frac{g(F_t)}{R_{f,t+1}} + \int_0^{F_t} g''(K) \, \mathrm{put}_t(K) \, dK + \int_{F_t}^{\infty} g''(K) \, \mathrm{call}_t(K) \, dK, \qquad 17.$$

where I write $\mathrm{call}_t(K)$ for the time t price of a European call on S_{t+1} that expires at time $t+1$ with strike K and $\mathrm{put}_t(K)$ for the corresponding put option. This form of the result is due to Carr and Madan (1998).

To find an expression for risk-neutral variance, we consider the case $g(K) = K^2$ (and recall that $F_t = \mathbb{E}_t^* \, S_{t+1} = S_t R_{f,t+1}$ and the return on the asset is $R_{t+1} = S_{t+1}/S_t$). We then have

$$\mathrm{var}_t^* \, R_{t+1} = \frac{2R_{f,t+1}}{S_t^2} \left\{ \int_0^{F_t} \mathrm{put}_t(K) \, dK + \int_{F_t}^{\infty} \mathrm{call}_t(K) \, dK \right\}. \qquad 18.$$

Martin (2017) defined the SVIX index via the formula

$$\mathrm{SVIX}_t^2 = \mathrm{var}_t^* \, \frac{R_{t+1}}{R_{f,t+1}} = \frac{2}{R_{f,t+1} S_t^2} \left\{ \int_0^{F_t} \mathrm{put}_t(K) \, dK + \int_{F_t}^{\infty} \mathrm{call}_t(K) \, dK \right\}, \qquad 19.$$

so that the equity premium perceived by the log investor satisfies

$$\mathbb{E}_t \, R_{t+1} - R_{f,t+1} = R_{f,t+1} \, \mathrm{SVIX}_t^2 \, . \qquad 20.$$

Equation 20 makes it possible to measure the log investor's perceived risk premium in real time via option prices. Knox and Vissing-Jorgensen (2024) and Knox et al. (2024) exploit this feature to interpret market responses to news events.

Table 1 reports results of regressions of realized returns onto SVIX_t^2, with Newey–West standard errors reported in square brackets. The left panel shows results over a sample period running from January 1996 to December 2022, extending the sample period studied in Martin (2017). Equation 20 predicts that we should find $\alpha = 0$ and $\beta = 1$. These predictions are not rejected by the data; and at the 6- and 12-month horizons (though not the 1- and 3-month horizons) we can reject the hypothesis that $\beta = 0$ at the 5% level.

As the data of Martin (2017) ended in January 2012, the right panel of Table 1 conducts an out-of-sample test by reporting coefficients estimated over the later part of the sample period, from February 2012 to December 2022. Over this shorter period, SVIX_t^2 is highly significant at the shorter forecasting horizons, marginally significant at the 6-month horizon (with a p-value of 0.066) and not significant at conventional levels at the 12-month horizon (with a p-value of 0.18).[13]

The point estimates of β in Table 1 are larger than one at all horizons, and for both sample periods; and they are statistically significantly larger than one at the shorter horizons over the recent sample period. This raises the possibility that equation 20 understates the true equity premium. Back et al. (2022) argue that this is indeed the case; I return to this issue in Section 3.2.

SVIX, VIX, CVOL . . . and SVIX

The SVIX index can be compared with the VIX index, which is defined by the formula

$$\mathrm{VIX}_t^2 = 2R_{f,t+1}\left\{\int_0^{F_t} \frac{1}{K^2}\,\mathrm{put}_t(K)\,dK + \int_{F_t}^{\infty} \frac{1}{K^2}\,\mathrm{call}_t(K)\,dK\right\}. \qquad 21.$$

Due to the weighting function $1/K^2$ inside the integrals, the VIX index places more weight on deep-out-of-the-money put options than SVIX, and less weight on deep-out-of-the-money calls; empirically, it spikes even more dramatically than SVIX at times of crisis. It is easy to check, using equation 17, that $\mathrm{VIX}_t^2 = 2\,\mathrm{L}_t^*\left(\frac{R_{t+1}}{R_{f,t+1}}\right)$, where the risk-neutral entropy operator, L_t^*, is defined by $\mathrm{L}_t^*(X) = \log \mathbb{E}_t^* X - \mathbb{E}_t^* \log X$. Martin (2011) and Martin (2017) discuss the relationship between SVIX, VIX, and variance swap markets in more detail; note, however, that Martin (2011) uses a definition of SVIX_t^2 that differs from the definition given in equation 19 by a factor of $R_{f,t+1}^2$.

Recently the Chicago Mercantile Exchange has launched a suite of volatility indices ("CVOL") based on the SVIX formula given in equation 19. These indices measure the risk-neutral variances of a range of asset classes.

Lastly, note that an ETF has recently been introduced under the name SVIX. This is an "inverse" or "short" ETF that dynamically shorts VIX futures contracts, so that it typically moves in the *opposite* direction to VIX and to SVIX as defined above.

[13]The short-horizon results differ sharply when the earlier data is included because of the period from October 2008 to March 2009. SVIX exploded in October and November 2008—predicting very high returns according to the theory of this section—but in the event the market continued to decline until March 2009 before rebounding.

2.1.2. The market's expected log return.

Equation 13 straightforwardly supplies the log investor's expectations about other functions of the market return. For example, the expected *log* return represents a useful measure of risk-adjusted returns. Indeed, from the log investor's point of view, the expected log return is precisely the right measure of investment opportunities, as it represents his or her expected utility if current wealth is normalized to one.

Expected log returns are also the natural quantity of interest when working with log-linear approximate identities as in Campbell and Shiller (1988). For this reason Gao and Martin (2021) use equation 13 to infer the log investor's expected log return:

$$\mathbb{E}_t \log R_{t+1} = \log R_{f,t+1} + \frac{1}{S_t}\left\{\int_0^{F_t} \frac{\text{put}_t(K)}{K}\,dK + \int_{F_t}^{\infty} \frac{\text{call}_t(K)}{K}\,dK\right\}. \qquad 22.$$

Another illustration of the convenience of log returns is provided by Gandhi et al. (2023), who seek to measure "forward return expectations". For example, $\mathbb{E}_t \log R_{t+1\to t+2} = \mathbb{E}_t \log R_{t\to t+2} - \mathbb{E}_t \log R_{t\to t+1}$, so that forward expectations from $t+1$ to $t+2$ can be inferred from one- and two-period expected log returns; and, under the log investor assumption, these can be evaluated using options maturing in, respectively, one and two periods in the formula given in equation 22.

2.1.3. The autocorrelation of the market.

The corresponding relationship for simple returns does not decompose in this convenient way: as $\mathbb{E}_t\, R_{t+1\to t+2} = \mathbb{E}_t \frac{R_{t\to t+2}}{R_{t\to t+1}} \neq \frac{\mathbb{E}_t\, R_{t\to t+2}}{\mathbb{E}_t\, R_{t\to t+1}}$, the relationship between spot and forward simple returns is sensitive to the autocorrelation in returns. As it happens, Martin (2021) shows that it is possible to infer the log investor's perceived autocorrelation of the stock market if *forward-start* index option prices are observed; but unfortunately these are rather exotic derivatives and the market for them is not very liquid.

2.1.4. The probability of a market crash.

By setting $X_{t+1} = \mathbf{1}_{R_{t+1}<x}$ in equation 13, we find that

$$\mathbb{P}_t\left(R_{t+1} < x\right) = \frac{1}{R_{f,t+1}}\,\mathbb{E}_t^*\left(R_{t+1}\mathbf{1}_{R_{t+1}<x}\right). \qquad 23.$$

$\mathbf{1}_{R_{t+1}<x}$: an indicator function that takes the value 1 if $R_{t+1} < x$ and 0 otherwise

Martin (2017) shows that the quantity on the right-hand side of equation 23 can be inferred from index option prices:

$$\mathbb{P}_t\left(R_{t+1} < x\right) = x\left[\text{put}_t'(xS_t) - \frac{\text{put}_t(xS_t)}{xS_t}\right]. \qquad 24.$$

Here $\text{put}_t'(xS_t)$ is the slope of the put price curve, plotted as a function of strike, K, at the point $K = xS_t$.

Goetzmann et al. (2024) use the formula (24) as a measure of the market-implied probability of a crash, and compare it to survey expectations of crashes.

2.1.5. The variance risk premium.

We can calculate the log investor's perceived forward looking true market variance, $\text{var}_t\, R_{t+1}$, by setting $X_{t+1} = R_{t+1}^2$ in equation 13. The variance risk premium as perceived by the log investor is determined by the relationship between risk-neutral variance and risk-neutral skewness, via the formula

$$\text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} - \text{var}_t \frac{R_{t+1}}{R_{f,t+1}} = \left(\text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}}\right)^2 - \mathbb{E}_t^*\left[\left(\frac{R_{t+1}}{R_{f,t+1}} - 1\right)^3\right]. \qquad 25.$$

Once again, the risk-neutral quantities can be calculated from option prices using equation 17. Essentially this formula is derived by Hsieh et al. (2024) and proposed as an index of the variance risk premium.

Martin (2017, Online Appendix) carries out this exercise—though without explicitly stating the above equation—and reports time series of true and risk-neutral volatility over the period 1996–2012. Over this period, risk-neutral volatility typically exceeds true volatility by on the order of 1 to 2 percentage points (annualised) at the 1-month and 1-year horizons; but at the height of the subprime crisis, the gap between the two spikes to around 6 percentage points at the 1-month horizon and 4 percentage points at the 1-year horizon.

Empirically, the variance risk premium (as calculated in equation 25) is always positive. In contrast, the approach of Bollerslev et al. (2009), which uses recent realized variance to proxy for forward-looking variance, delivers the puzzling finding that the variance risk premium sometimes spikes *downwards* and below zero at times of market stress.

2.2. Other assets

The same logic that led to equation 15 implies, under the log investor assumption, that the expected return on an arbitrary asset i must satisfy

$$\mathbb{E}_t\, R_{i,t+1} - R_{f,t+1} = \frac{1}{R_{f,t+1}} \operatorname{cov}_t^* \left(R_{i,t+1}, R_{t+1}\right) . \qquad 26.$$

At first sight, equation 26 takes a familiar form: it says that the asset's expected excess return should be proportional to its covariance with the market return, as in the CAPM. Here, though, the relevant quantity is the conditional *risk-neutral* covariance.

In principle, this has the advantage exploited in the previous subsection: one can hope to measure risk-neutral covariance directly from asset prices without further assumptions. But whereas it is easy to use option prices to pin down risk-neutral expectations of functions of a single variable, as in equation 17, one cannot in general hope to determine risk-neutral expectations of arbitrary functions of two or more variables given the assets that are traded in practice (Martin 2018). Vanilla options provide information about the univariate risk-neutral distributions of the assets on which they are written, but they do not identify the *joint* risk-neutral distribution.

2.2.1. Currencies. In the case of currencies, however, a minor miracle occurs: a contract that reveals the risk-neutral covariance between (say) the yen and the S&P 500 index happens to be traded.

To apply equation 26 to currencies, we need to interpret $R_{i,t+1}$ as the return on a currency trade. If the time t price of a unit of foreign currency is $\$e_{i,t}$, then at time t we can take \$1 and convert it to $1/e_{i,t}$ units of foreign currency. Having done so, we invest it until time $t+1$ at the foreign-currency interest rate, $R^i_{f,t+1}$, then convert back to dollars at time $t+1$. The dollar return on the currency trade is therefore $\frac{e_{i,t+1}}{e_{i,t}} R^i_{f,t+1}$. Substituting this quantity for $R_{i,t+1}$ in equation 26 and rearranging, we have

$$\mathbb{E}_t\, \frac{e_{i,t+1}}{e_{i,t}} - 1 = \underbrace{\frac{R_{f,t+1}}{R^i_{f,t+1}} - 1}_{\mathrm{IRD}_{i,t}} + \underbrace{\frac{1}{R_{f,t+1}} \operatorname{cov}_t^* \left(\frac{e_{i,t+1}}{e_{i,t}}, R_{t+1}\right)}_{\mathrm{QRP}_{i,t}} . \qquad 27.$$

Equation 27 expresses expected currency movement as the sum of two terms. The first is the interest-rate differential, $\mathrm{IRD}_{i,t}$. This is the expected currency appreciation according

Table 2: Forecasting currency movements. Monthly data, 2009.12–2017.10

Pooled panel regression: $\frac{e_{i,t+1}}{e_{i,t}} - 1 = \alpha + \beta \mathrm{QRP}_{i,t} + \gamma \mathrm{IRD}_{i,t} + \varepsilon_{i,t+1}$				
horizon	α	β	γ	R^2 (%)
24mo	−0.048 [0.020]	3.394 [1.726]	1.769 [1.045]	16.01
24mo	−0.030 [0.014]		0.168 [0.651]	0.16
Panel regression with fixed effects: $\frac{e_{i,t+1}}{e_{i,t}} - 1 = \alpha_i + \beta \mathrm{QRP}_{i,t} + \gamma \mathrm{IRD}_{i,t} + \varepsilon_{i,t+1}$				
horizon		β	γ	R^2 (%)
24mo		5.456 [2.047]	1.717 [1.414]	20.56
24mo			−0.363 [1.007]	0.20

Bootstrapped standard errors are reported in square brackets. Results are from Table 5 of Kremens and Martin (2019).

to the theory of uncovered interest parity (UIP), in which exchange rates are expected to appreciate or depreciate in such a way that all currency trades earn identical expected returns. This simplistic prediction neglects the impact of risk, which is captured above in the second term. From the log investor's point of view, a currency's risk premium is revealed by its risk-neutral covariance with the market. Kremens and Martin (2019) derive equation 27 and show that the risk-neutral covariance term is revealed by comparing the forward price of the market to the quanto forward price of the market.

Quanto contracts

Someone who goes long a conventional forward contract on the market commits, at time t, to pay the known amount F_t at time $t+1$ in exchange for the then prevailing level of the market, P_{t+1}. Here both F_t and P_{t+1} are denominated in dollars. By contrast, someone who goes long a currency-i quanto forward contract on the market commits, at time t, to pay the known amount $Q_{i,t}$ units of currency i at time $t+1$ in exchange for P_{t+1} *units of currency* i. This contract is sensitive to the correlation between currency i and the market: if, say, currency i depreciates catastrophically whenever the market does well, then all else equal this makes the quanto contract unattractive, so that $Q_{i,t}$ will have to be small. Specifically, Kremens and Martin (2019) show that

$$\frac{Q_{i,t} - F_t}{R^i_{f,t+1} P_t} = \frac{1}{R_{f,t+1}} \operatorname{cov}^*_t \left(\frac{e_{i,t+1}}{e_{i,t}}, R_{t+1} \right), \qquad 28.$$

and refer to the quantity on the left-hand side as the quanto-implied risk premium, $\mathrm{QRP}_{i,t}$.

Table 2 reports the results of regressions of realized currency appreciation onto QRP and IRD, or onto IRD alone, taken from Kremens and Martin (2019). The forecasting horizon is two years, to match the horizon of observable quanto contracts. According to equation 27 we should expect to find a zero intercept, and estimated coefficients $\beta = 1$

on QRP and $\gamma = 1$ on IRD. In a pooled panel regression, the results do not reject this hypothesis, and the coefficient on QRP is significantly different from zero. The inclusion of QRP in the panel regressions increases R^2 by two orders of magnitude relative to what IRD achieves on its own; and the estimated coefficient on IRD moves in the right direction (i.e., towards 1) when QRP is included, though it is not significantly different from zero in either specification.

The bottom panel of the table reports broadly consistent results when currency fixed effects are included, but the coefficient on QRP, which remains statistically significant, is now also significantly greater than one, suggesting that the log investor's view understates the magnitude of currency risk premia. (I discuss this fact further, and explain the motivation for including currency fixed effects, in Section 3.) Once again, including QRP increases R^2 by two orders of magnitude.

Kremens et al. (2024) connect this theory to the data in a different way, showing, for six high-income currencies, that expected two-year currency movements drawn from surveys of professional forecasters successfully forecast outcomes and correlate strongly with QRP and a small number of other macro-finance variables (notably the real exchange rate and current account-GDP ratio).

2.2.2. Individual stocks. To apply equation 26 to individual stocks, we would like to be able to observe the risk-neutral covariance between stock i and the S&P 500 index. This would be feasible if, say, there were a liquid market in "outperformance options" (that is, options on $R_{i,t+1} - R_{t+1}$): as $\text{cov}_t^*(R_{i,t+1}, R_{t+1}) = \frac{1}{2}\left[\text{var}_t^* R_{i,t+1} + \text{var}_t^* R_{t+1} - \text{var}_t^*\left(R_{i,t+1} - R_{t+1}\right)\right]$, we could infer risk-neutral covariance by observing stock i options, index options, and outperformance options. But there is no such market at present.

Martin and Wagner (2019) therefore take another tack, exploiting the fact that typical stocks have betas close to one. Note first that equation 26 can be rewritten

$$\mathbb{E}_t \frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \beta_{i,t}^* \,\text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}}. \qquad 29.$$

Here $\beta_{i,t}^*$ is the risk-neutral beta of stock i with respect to the market, associated with the decomposition

$$\frac{R_{i,t+1}}{R_{f,t+1}} = \alpha_{i,t}^* + \beta_{i,t}^* \frac{R_{t+1}}{R_{f,t+1}} + \varepsilon_{i,t+1}^* \qquad 30.$$

where

$$\beta_{i,t}^* = \frac{\text{cov}_t^*\left(R_{i,t+1}, R_{t+1}\right)}{\text{var}_t^* R_{t+1}} \qquad 31.$$

$$\mathbb{E}_t^* \,\varepsilon_{i,t+1}^* = 0 \qquad 32.$$

$$\text{cov}_t^*\left(\varepsilon_{i,t+1}^*, R_{t+1}\right) = 0. \qquad 33.$$

Equations 30–33 can be viewed as characterizing a "risk-neutral regression" of stock i's return onto the market return. They imply that

$$\text{var}_t^* \frac{R_{i,t+1}}{R_{f,t+1}} = \beta_{i,t}^{*2} \,\text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} + \text{var}_t^* \,\varepsilon_{i,t+1}^*. \qquad 34.$$

As already noted, the quantity on the right-hand side of equation 29 cannot be directly observed from vanilla option prices. But it is related to the quantity on the right-hand side of equation 34, and hence to $\text{var}_t^* R_{i,t+1}$—and this *is* observable given option prices on stock i. To make this relationship precise, we use the linearization $\beta_{i,t}^{*2} \approx 2\beta_{i,t}^* - 1$ in equation 34 to find, after some rearrangement,

$$\beta_{i,t}^* \text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} = \frac{1}{2} \text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} + \frac{1}{2} \text{var}_t^* \frac{R_{i,t+1}}{R_{f,t+1}} - \frac{1}{2} \text{var}_t^* \varepsilon_{i,t+1}^*. \tag{35.}$$

It follows from equation 29 that

$$\mathbb{E}_t \frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \frac{1}{2} \text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} + \frac{1}{2} \text{var}_t^* \frac{R_{i,t+1}}{R_{f,t+1}} - \frac{1}{2} \text{var}_t^* \varepsilon_{i,t+1}^*. \tag{36.}$$

Multiplying by value weights, $w_{i,t}$, and summing over i,

$$\mathbb{E}_t \frac{R_{t+1}}{R_{f,t+1}} - 1 = \frac{1}{2} \text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} + \frac{1}{2} \sum_i w_{i,t} \text{var}_t^* \frac{R_{i,t+1}}{R_{f,t+1}} - \frac{1}{2} \sum_i w_{i,t} \text{var}_t^* \varepsilon_{i,t+1}^*. \tag{37.}$$

Subtracting equation 37 from equation 36 and using the fact that $\mathbb{E}_t \frac{R_{t+1}}{R_{f,t+1}} - 1 = \text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}}$, we have

$$\begin{aligned}\mathbb{E}_t \frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \text{var}_t^* \frac{R_{t+1}}{R_{f,t+1}} &+ \frac{1}{2} \left\{ \text{var}_t^* \frac{R_{i,t+1}}{R_{f,t+1}} - \sum_j w_{j,t} \text{var}_t^* \frac{R_{j,t+1}}{R_{f,t+1}} \right\} \\ &- \frac{1}{2} \left\{ \text{var}_t^* \varepsilon_{i,t+1}^* - \sum_j w_{j,t} \text{var}_t^* \varepsilon_{j,t+1}^* \right\}.\end{aligned} \tag{38.}$$

The third term on the right-hand side of equation 38 is zero on value-weighted average. Martin and Wagner (2019) make the econometrically convenient assumption that it is constant over time so that it can be replaced by a fixed effect α_i. By analogy with the definition of SVIX given in equation 19, they define $\text{SVIX}_{i,t}^2 = \text{var}_t^* \frac{R_{i,t+1}}{R_{f,t+1}}$ and $\overline{\text{SVIX}}_t^2 = \sum_j w_{j,t} \text{var}_t^* \frac{R_{j,t+1}}{R_{f,t+1}}$: these quantities can be inferred using options on individual stocks in the formula given in equation 19. The end result is the formula

$$\mathbb{E}_t \frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \alpha_i + \text{SVIX}_t^2 + \frac{1}{2} \left(\text{SVIX}_{i,t}^2 - \overline{\text{SVIX}}_t^2 \right). \tag{39.}$$

The fixed effects α_i are zero on weighted average, so if they are constant across i then they must all equal zero. In this case

$$\mathbb{E}_t \frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \text{SVIX}_t^2 + \frac{1}{2} \left(\text{SVIX}_{i,t}^2 - \overline{\text{SVIX}}_t^2 \right). \tag{40.}$$

Table 3 reports results from Martin and Wagner (2019), who test these equations by regressing realizations onto predictions. The top panel shows pooled results, testing the more aggressive prediction in equation 40 (according to which $\alpha = 0$, $\beta = 1$, and $\gamma = 1/2$); the bottom panel shows results with fixed effects, testing the prediction of equation 39 (according to which the time series average of the value-weighted sum of fixed effects, $\sum_i w_i \alpha_i$, equals zero, $\beta = 1$, and $\gamma = 1/2$). Fixed effects appear to matter: when they are included, the predictions of equation 39 are not rejected at any horizon, and the null that the coefficients are zero is strongly rejected at horizons of 6, 12, and 24 months.

Table 3: Forecasting individual stock returns. S&P 500 firms. Monthly, 1996.01–2014.10

Pooled panel regressions: $\frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \alpha + \beta\,\text{SVIX}_t^2 + \gamma\left(\text{SVIX}_{i,t}^2 - \overline{\text{SVIX}}_t^2\right) + \varepsilon_{i,t+1}$				
horizon	α	β	γ	R^2 (%)
1mo	0.057 [0.074]	0.743 [2.311]	0.214 [0.296]	0.096
6mo	−0.038 [0.059]	3.483 [1.569]	0.463 [0.320]	3.218
12mo	−0.021 [0.071]	3.032 [1.608]	0.512 [0.318]	4.423
24mo	−0.054 [0.076]	3.933 [1.792]	0.324 [0.200]	5.989
Panel regressions with fixed effects: $\frac{R_{i,t+1}}{R_{f,t+1}} - 1 = \alpha_i + \beta\,\text{SVIX}_t^2 + \gamma\left(\text{SVIX}_{i,t}^2 - \overline{\text{SVIX}}_t^2\right) + \varepsilon_{i,t+1}$				
horizon	$\sum w_i \alpha_i$	β	γ	R^2 (%)
1mo	0.080 [0.072]	0.603 [2.298]	0.491 [0.325]	0.650
6mo	−0.008 [0.055]	3.161 [1.475]	0.892 [0.336]	10.356
12mo	0.012 [0.070]	2.612 [1.493]	0.938 [0.308]	17.129
24mo	−0.026 [0.079]	3.478 [1.681]	0.665 [0.205]	24.266

Bootstrapped standard errors are reported in square brackets. Results are from Tables IV and V of Martin and Wagner (2019).

2.2.3. Dividends. Throughout the paper, I have neglected the distinction between total returns and capital gains, effectively treating the contribution of dividends to returns as negligible. This assumption is forced because at present options on US stocks and indices are typically written on ex-dividend prices, rather than on total returns.[14] The assumption is tolerable over shorter horizons (or for non-divided-paying assets such as currencies) but would become problematic once the forecasting horizon rises substantially above a year or two, as I will discuss further in Section 2.2.4.

There is, however, a growing market in claims on the dividends of the aggregate market paid over a given year. Gormsen et al. (2021) use them to understand expectations about aggregate dividends, building on earlier work of Gormsen and Koijen (2020). I will write P_t^d for the price of a claim to dividends over the period from t to $t+1$, and $R_{t+1}^d = D_{t+1}/P_t^d$ for the return on this claim. From the perspective of the log investor, we then have

$$\mathbb{E}_t\, R_{t+1}^d - R_{f,t+1} = \frac{1}{R_{f,t+1}} \operatorname{cov}_t^*\left(R_{t+1}^d, R_{t+1}\right). \qquad 41.$$

If we separately observed options on the total return, on the market capital gain, and on dividends, we could determine $\operatorname{var}_t^*(S_{t+1} + D_{t+1})$, $\operatorname{var}_t^* S_{t+1}$, and $\operatorname{var}_t^* D_{t+1}$. Together these would pin down $\operatorname{cov}_t^*(D_{t+1}, S_{t+1})$, and it would then be possible to calculate the covariance

[14] There is no obvious reason for this to be the case other than market convention—and options on a total return would be easier for market-maker to hedge than options on a capital gain—so things may change in future.

term in equation 41 without further assumptions.

In the absence of such data, Gormsen et al. (2021) make an observation that may be useful in other contexts. They note that if two gross returns—in this case, R_{t+1}^d and R_{t+1}—are jointly lognormal, then their risk-neutral correlation equals their true correlation. In this case, we can decompose the risk-neutral covariance as the product of true correlation and two risk-neutral volatility terms:

$$\operatorname{cov}_t^*\left(R_{t+1}^d, R_{t+1}\right) = \rho_t \sigma_t^*\left(R_{t+1}^d\right)\sigma_t^*\left(R_{t+1}\right). \qquad 42.$$

The two risk-neutral volatility terms are observable from traded option prices, so this equation could be implemented either by assuming that ρ_t takes a particular value, or lies in some range, or by using realized correlation over a recent time period to proxy for ρ_t.

2.2.4. Interest rates. To illustrate how the log investor perspective can suggest directions that future research might take, we can ask which asset prices would, in principle, reveal the log investor's expectations of future interest rates.

Consider two alternative ways of investing money in bond markets. One is a rolling investment at a floating short rate of interest. This strategy is riskless when considered one period at a time, but is exposed to variation in interest rates over the long run. I write $R_{c,u\to v}$ to denote the gross return, from time u to time v, on a "cash" strategy that repeatedly invests at the short (i.e., one-period) interest rate: thus $R_{c,t\to T} = R_{f,t+1}R_{f,t+2}\cdots R_{f,T}$. The other strategy is to invest at a fixed long rate of interest. I write $R_{b,u\to v}$ to denote the gross fixed riskless rate that can be locked in between time u and time v: this is the gross return, between time u and time v, on a zero-coupon bond, which can be determined from the $(v-u)$-period yield at time u. (Note that $R_{b,t\to t+1} = R_{f,t+1}$.)

The difference between $\mathbb{E}_t\, R_{c,t\to T}$ and $R_{b,t\to T}$ (the latter being a known constant at time t) is a measure of the *expected future path of interest rates*. To determine the equilibrium value of this quantity in the mind of the log investor, we exploit a relationship between futures and forward prices derived by Cox et al. (1981).

Consider an index futures contract with settlement date T. On date t, the futures price is G_t. By definition of the contract, at date T the futures price settles at the then prevailing index price: $G_T = S_T$. (As noted above, we assume that the index is quoted as a total return, so $R_{t\to T} = S_T/S_t$.) No money changes hands at initiation of a trade, on (say) day t. The next day—day $t+1$—the long side of the trade receives $G_{t+1} - G_t$ from the short side. As it was costless to enter the trade, it must be the case that $\mathbb{E}_t\left[M_{t+1}\left(G_{t+1} - G_t\right)\right] = 0$. As this relationship holds for all $t < T$, we can work backwards to conclude that

$$G_t = \mathbb{E}_t\left[M_{t+1}\cdots M_T R_{c,t\to T} S_T\right]. \qquad 43.$$

We suppose now that the log investor is maximizing expected utility of wealth at some future date T; as before, today is date t.[15] I write $R_{t\to T}$ for the gross return on the market from time t to time T. From the log investor's perspective, $\mathbb{E}_t\, R_{c,t\to T}$ equals the price of a claim to $R_{c,t\to T}R_{t\to T}$, by equation 13: so, from equation 43, our measure of the expected

[15]This is consistent with what we did before: as long-horizon log returns decompose separably into a sum of per-period log returns, this log investor will continue to ensure that he or she is at an optimum for the problem depicted in equation 11.

future path of interest rates is

$$\mathbb{E}_t\, R_{c,t\to T} - R_{b,t\to T} = \frac{G_t}{S_t} - R_{b,t\to T} = \frac{G_t - F_t}{S_t} \qquad 44.$$

where F_t and G_t are, respectively, the forward and futures prices of the index to time T, calculated at time t, and S_t is the spot price.[16] If the futures price exceeds the forward price, this signals that there is a positive risk-neutral correlation between short rates and the market, so that the cash trade underperforms when the market underperforms; as a result, the cash trade must earn a risk premium.

As noted above, this subsection is illustrative of a direction that future research might take. For this approach to deliver interesting predictions, we would need to look at reasonably long horizons, with T on the order of five years or more. At present, however, there are no liquid long-dated index futures contracts on US stock markets.[17]

3. A GENERAL FRAMEWORK

The results of the last section all followed from equation 13, which relies on the log investor assumption. That equation can be generalized to the following *identity*, which requires no assumptions on the form of the SDF:

$$\mathbb{E}_t\, X_{t+1} = \mathbb{E}_t^*\, X_{t+1} + \frac{1}{R_{f,t+1}} \operatorname{cov}_t^* \left(X_{t+1}, R_{t+1}\right) - \operatorname{cov}_t \left(M_{t+1} R_{t+1}, X_{t+1}\right). \qquad 45.$$

Essentially this identity (specialized to the case in which $X_{t+1} = R^i_{f,t+1} e_{i,t+1}/e_{i,t}$ is the return on a currency trade) was derived by Kremens and Martin (2019). It holds for arbitrary X_{t+1} and an arbitrary gross return R_{t+1}.

If R_{t+1} is chosen to equal the return on the market, the second of the two covariance terms drops out entirely in the log investor case as $M_{t+1}R_{t+1} = 1$. Alternatively, if R_{t+1} is chosen to equal the riskless rate, the identity 45 reduces to the more familiar identity

$$\mathbb{E}_t\, X_{t+1} = \mathbb{E}_t^*\, X_{t+1} - \operatorname{cov}_t \left(M_{t+1} R_{f,t+1}, X_{t+1}\right) \qquad 46.$$

which, when applied to a return, $X_{t+1} = R_{i,t+1}$, shows that an asset's risk premium is determined by the covariance between its return and the SDF. (Recall that $\mathbb{E}_t^*\, R_{i,t+1} = R_{f,t+1}$ for any return.)

The identities 45 and 46 each relate the expectation of interest, $\mathbb{E}_t\, X_{t+1}$, to its risk-neutral counterpart—which is, in principle, observable directly from asset prices—and to covariance terms.

Identity 46 tells us that for the risk-neutral expectation $\mathbb{E}_t^*\, X_{t+1}$ to be a useful measure of $\mathbb{E}_t\, X_{t+1}$, X_{t+1} must be approximately conditionally uncorrelated with the SDF. This is the (explicit or implicit) assumption when CDS rates are used as surrogates for expected

[16] I continue to assume that the index is quoted on a total return basis. If not, the forward and futures prices are each adjusted to account for dividends not received. As equation 44 exploits the *gap* between the two, however, we can expect some cancellation so that the importance of neglecting dividends may be relatively minor even at longer horizons.

[17] The CME has introduced an Adjusted Interest Rate (AIR) Total Return futures contract, but the contract is explicitly designed not to have the exposure to future short interest rates that a conventional index futures contract has, and which the above analysis exploits, as in equation 43.

default rates, breakeven inflation for expected inflation, forward rates for expected future interest rates, and so on. Unfortunately, the assumption is implausible in these cases, and in most other cases of interest to observers of financial markets.

In contrast, the identity 45 includes a further risk-neutral quantity, $\text{cov}_t^*(X_{t+1}, R_{t+1})$. If the identity is applied with R_{t+1} equal to the return on the market, this quantity is proportional to a "risk-neutral market beta": it is a risk adjustment that is potentially directly observable from asset prices, as in Subsection 2.2.1. Moreover, as the SDF and market return typically move in opposite directions, it is then reasonable to hope that the remaining "nuisance" covariance term, $\text{cov}_t(M_{t+1}R_{t+1}, X_{t+1})$, is smaller than the corresponding term, $\text{cov}_t(M_{t+1}R_{f,t+1}, X_{t+1})$, in identity 46.

The remainder of this section discusses various approaches proposed in the literature to handle the term $\text{cov}_t(M_{t+1}R_{t+1}, X_{t+1})$. Under the log investor assumption, it is literally zero, as already noted; and this continues to be true with Epstein and Zin (1989) and Weil (1990) preferences with unit risk aversion and arbitrary coefficient of intertemporal substitution; or, more generally, if $M_{t+1}R_{t+1}$ is uncorrelated with X_{t+1}.

3.1. A reduced-form approach

A pragmatic reaction is that we should include other explanatory variables, in addition to $\frac{1}{R_{f,t+1}}\text{cov}_t^*(X_{t+1}, R_{t+1})$, to proxy for $-\text{cov}_t(M_{t+1}R_{t+1}, X_{t+1})$. As equation 45 is an identity, this approach is free of assumptions. Analogously, one can think of the predictor variables in the conventional reduced-form approach to return forecasting as capturing the term $-\text{cov}_t(M_{t+1}R_{f,t+1}, X_{t+1})$ in identity 46.

When identity 45 is applied with R_{t+1} equal to the return on a broad market index such as the S&P 500, it is highly plausible that R_{t+1} offsets some of the movement in M_{t+1}, so that $M_{t+1}R_{t+1}$ comoves less with X_{t+1} than $M_{t+1}R_{f,t+1}$ does. Identity 45 therefore has the advantage, relative to identity 46, that these other explanatory variables have less to explain than they do in the conventional approach.

As an illustration of this approach, the fixed effects included in Tables 2 and 3 can be thought of as capturing the cross-sectional, time-invariant component of the covariance term in applications to currencies and to stocks. Moreover, Kremens and Martin (2019) find that while QRP is a highly significant predictor of currency movements in the time series and cross-section, as shown in Table 2, other variables, notably the real exchange rate (Dahlquist and Pénasse 2022) and a dollar factor (the average forward discount of Lustig et al. (2014)) also enter significantly into currency forecasting regressions and substantially increase R^2 above what QRP achieves on its own.

3.2. Lower bounds

In some circumstances, it is possible to sign the second of the covariance terms in identity 45. Working in the case $X_{t+1} = R_{t+1}$, Martin (2017) shows that the *negative correlation condition (NCC)* $\text{cov}_t(M_{t+1}R_{t+1}, R_{t+1}) \leq 0$ holds from the perspective of an investor who holds the market if risk aversion is at least one, or alternatively under various conditions that cover leading macro-finance models such as Campbell and Cochrane (1999), Bansal and Yaron (2004), Wachter (2013), Bansal et al. (2014), and Campbell et al. (2018). Under the NCC, we then have the lower bound

$$\mathbb{E}_t R_{t+1} - R_{f,t+1} \geq \frac{1}{R_{f,t+1}} \text{var}_t^* R_{t+1}, \qquad 47.$$

and this bound is valid under considerably weaker assumptions than were required for equality to hold, as in equation 15.

The lower bound (47) is extremely volatile, right-skewed, and fat-tailed, exhibiting sharp peaks that die away fairly rapidly. As the peaks are far larger than reasonable measures of the unconditional equity premium, Martin (2017) emphasizes that these facts point to a spiky, volatile equity premium, and hence to a qualitatively different view than comes out of the literature that uses valuation ratios to forecast returns.[18]

Similarly, Gao and Martin (2021) exploit a lower bound on the expected log return based on the modified negative correlation condition (mNCC) $\operatorname{cov}_t\left(M_{t+1}R_{t+1}, \log R_{t+1}\right) \leq 0$. This holds under very similar conditions to the NCC, and in particular it holds in the macro-finance models mentioned above. When it holds, we have a lower bound $\mathbb{E}_t \log R_{t+1} \geq \frac{1}{R_{f,t+1}} \mathbb{E}_t^* \left(R_{t+1} \log R_{t+1}\right)$, and the right-hand side of this equation can be calculated from put and call prices using the formula 17.

Kadan and Tang (2020) determine conditions under which the lower bound (47) can be applied at the level of an individual stock. As the lower bound is given by the individual stock's risk-neutral variance—that is, by stock-level SVIX—it avoids the problem of measuring (or approximating, as in Section 2.2.2) the stock's risk-neutral covariance of the market. The cost of doing so is that the bound only applies for a stock i if the ratio $\operatorname{var}_t R_{i,t+1} / \operatorname{cov}_t(R_{i,t+1}, R_{t+1})$ exceeds the level of risk aversion.

3.3. Sharpening the lower bound

If we write $B_{t+1} = R_{t+1} - R_{f,t+1} - \frac{1}{R_{f,t+1}} \operatorname{var}_t^* R_{t+1}$, then the lower bound (47) asserts that $\mathbb{E}_t B_{t+1} \geq 0$. Back et al. (2022) test the validity of this claim, exploiting the fact that it implies an unconditional bound $\mathbb{E}\left(z_t B_{t+1}\right) \geq 0$ for any vector of positive conditioning variables z_t.

Using a range of variables drawn from Goyal and Welch (2008) in the conditioning vector z_t, they do not reject the hypothesis that the bound is valid—that is, that risk-neutral variance supplies a lower bound for the equity premium. But they also test the hypothesis that the bound is *tight* (i.e., holds with equality, as in the log investor case of Section 2.1.1), and this they can reject with moderate confidence (with finite-sample p-values of 3.6% and 8.3% at the one-month and one-year horizons over the period 1990–2020).

Taken at face value, this finding suggests at least two potential refinements of the log investor approach.

3.3.1. Allowing the log investor to trade more aggressively. Rather than assuming that the log investor holds the market, we can estimate the portfolio that a log investor would hold. The return on this portfolio, $R_{g,t+1}$, is referred to as the growth-optimal return (Kelly 1956, Long 1990), and its reciprocal is a stochastic discount factor if the first-order conditions in equation 12 have an interior solution. Then we have

$$\mathbb{E}_t R_{i,t+1} - R_{f,t+1} = \frac{1}{R_{f,t+1}} \operatorname{cov}_t^*\left(R_{i,t+1}, R_{g,t+1}\right) . \qquad 48.$$

This equation, which is the starting point of Martin and Wagner (2019), replaces the market return that appears in equation 26 with the growth-optimal return, whatever that may be.

[18]Early papers in this literature include Keim and Stambaugh (1986), Fama and French (1988), and Campbell and Shiller (1988).

Tetlock (2023) pursues this idea by estimating the growth-optimal return attainable by an investor who can trade the market and derivatives whose payoffs are the first, second, third, and fourth powers of the market's excess return. (As always, such contracts are observable from index option prices, by equation 17.) The key challenge is in the estimation of the portfolio weights (on the market and on the various power contracts) that determine the growth-optimal portfolio: Tetlock determines them by requiring the model to accurately match a measure of the variance risk premium. The estimated growth-optimal portfolio takes a levered position in the market that is largely funded by shorting the second and third power contracts. In other words, Tetlock argues that it would be optimal for a log investor to short volatility in order to lever up his or her market position, and finds that this estimated growth-optimal return forecasts market returns more accurately than does the lower bound (47), consistent with the results of Back et al. (2022).

The growth-optimal approach has the advantage that no assumptions need to be made about the identity of an investor who holds the market; but it has the disadvantage that the growth-optimal portfolio weights must be estimated. They are in general time-varying, and in principle may change suddenly at times of market turmoil.

3.3.2. Allowing the representative investor to be more risk-averse. As an alternative to thinking about the portfolio choices of the log investor, we can also take the perspective of an investor who has power utility with risk aversion γ over next period wealth, and who chooses to hold the market. The case $\gamma > 1$ allows for the possibility that a log investor would wish to trade more aggressively, as described in the previous subsection; but it has the advantage that there are no time-varying portfolio weights or other parameters to be estimated.

The logic that led to equation 12 implies that $M_{t+1} = \lambda_t R_{t+1}^{-\gamma}$, where λ_t is known at time t, and Martin (2017) shows that equation 13 is then replaced by

$$\mathbb{E}_t\, X_{t+1} = \frac{\mathbb{E}_t^*\left(R_{t+1}^{\gamma} X_{t+1}\right)}{\mathbb{E}_t^*\left(R_{t+1}^{\gamma}\right)} \quad \text{or} \quad \mathbb{E}_t\, X_{t+1} - \mathbb{E}_t^*\, X_{t+1} = \frac{\operatorname{cov}_t^*\left(R_{t+1}^{\gamma}, X_{t+1}\right)}{\mathbb{E}_t^*\, R_{t+1}^{\gamma}}. \qquad 49.$$

If X_{t+1} is a function of the return on the market itself, then the right-hand side is a ratio of risk-neutral expectations of functions of R_{t+1} that is easily evaluated using index options and the Breeden–Litzenberger approach, as in equation 17. For example, we can use equation 49 to write the market risk premium in a form comparable with the GBM case (equation 3). We have

$$\log \mathbb{E}_t\, \frac{R_{t+1}}{R_{f,t+1}} = \log \frac{\mathbb{E}_t^*\, R_{t+1}^{1+\gamma}}{\mathbb{E}_t^*\, R_{t+1}\, \mathbb{E}_t^*\, R_{t+1}^{\gamma}}. \qquad 50.$$

If, say, the investor who holds the market has risk aversion $\gamma = 2$, then the risk premium is determined by the first, second, and third risk-neutral moments of the market return (or, equivalently, by the risk-neutral skewness, variance, and mean of the market return). The right-hand side of equation 50 can be calculated using the following formula, which applies for arbitrary $\theta \in \mathbb{R}$ as yet another consequence of equation 17:

$$\mathbb{E}_t^*\, R_{t+1}^{\theta} = R_{f,t+1}^{\theta} + \frac{R_{f,t+1}\theta(\theta-1)}{S_t^{\theta}} \left\{ \int_0^{S_t R_{f,t+1}} K^{\theta-2}\, \text{put}_t(K)\, dK + \int_{S_t R_{f,t+1}}^{\infty} K^{\theta-2}\, \text{call}_t(K)\, dK \right\}. \qquad 51.$$

The left panel of Figure 1 reports the 1-month equity premium, calculated as in equation 50 and annualized by multiplying by 12, for γ equal to 1, 2, and 3. The right panel shows

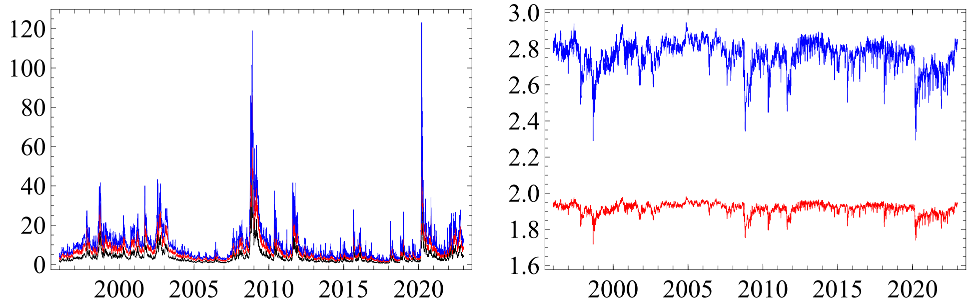

Figure 1: Left panel: Annualized 1-month equity premium calculated using equations 50 and 51, for γ equal to 1 (black), 2 (red), and 3 (blue). Right panel: The ratio of the implied equity premium to the log investor's equity premium, for γ equal to 2 (red) and 3 (blue).

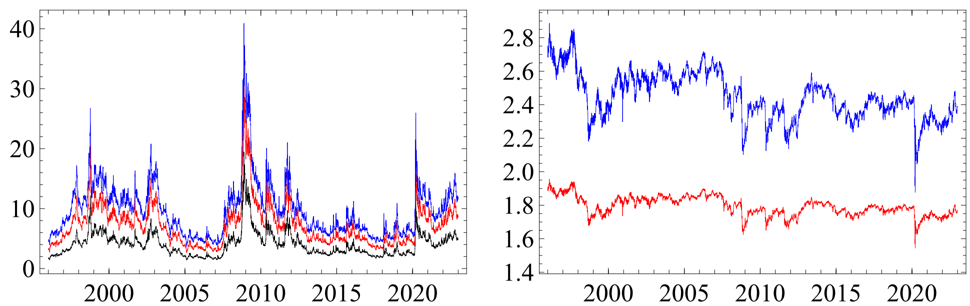

Figure 2: Left panel: 1-year equity premium calculated using equations 50 and 51, for γ equal to 1 (black), 2 (red), and 3 (blue). Right panel: The ratio of the implied equity premium to the log investor's equity premium, for γ equal to 2 (red) and 3 (blue).

the ratio of the implied equity premium to the log investor's equity premium for γ equal to 2 and 3. The implied equity premium grows more slowly with γ than would be predicted by a lognormal model. (If R_{t+1} were lognormal under the risk-neutral measure, with volatility σ, then the risk premium (50) would simplify to $\gamma\sigma^2$ so that the lines in the right panel would be constant at 2 and 3, respectively.)

Figure 2 repeats this exercise at the 1-year horizon. The non-linear scaling with γ is even more visible: the risk premium associated with $\gamma = 3$ is considerably less than three times as large as the log investor's perceived risk premium, and the ratio of the two tends to shrink in periods of high volatility.

More generally, once a functional form $M_{t+1} = f(R_{t+1})$ is specified,[19] one can calculate

$$\mathbb{E}_t \, g(R_{t+1}) = \mathbb{E}_t \left[M_{t+1} \frac{g(R_{t+1})}{f(R_{t+1})} \right] = \frac{1}{R_{f,t+1}} \, \mathbb{E}^*_t \left[\frac{g(R_{t+1})}{f(R_{t+1})} \right] \qquad 52.$$

and the quantity on the right-hand side can be determined from observable option prices using equation 17. Chabi-Yo and Loudis (2020) use this approach with $g(R_{t+1}) = R_{t+1}$ to estimate the equity premium.

[19]In fact, as $\mathbb{E}_t \left[M_{t+1} h(R_{t+1}) \right] = \mathbb{E}_t \left[\mathbb{E}_t \left(M_{t+1} \mid R_{t+1} \right) h(R_{t+1}) \right]$, all we need is that $\mathbb{E}_t \left(M_{t+1} \mid R_{t+1} \right)$, is a known function $f(R_{t+1})$ of the market return.

3.4. The correlation structure

If the variable we wish to forecast, X_{t+1}, is a function of quantities other than the market return, then the inference problem is challenging even with log utility, as discussed in Section 2.2. Quanto contracts reveal risk-neutral covariances of the market with currency movements, so we can infer the log investor's expectations about currency movements. But we do not at present observe contracts that reveal, say, the risk-neutral covariance of the market with inflation, so we cannot infer inflation expectations.

The problem in this example is pervasive—and fundamental, because of the importance of covariances throughout financial economics. Option prices are observable on a wide range of underlying payoffs—equity indices, individual stocks, currencies, interest rates, bond prices, inflation, and so on—and they reveal the associated univariate (or marginal) risk-neutral distributions. But vanilla options do not reveal the joint risk-neutral distributions we need to observe to implement equation 26 or 49 (Martin 2018).

As noted in Section 2.2.2, this fact provides a motivation for the introduction of new markets. If, say, we observed options on the outperformance of the market relative to a 10-year bond, $R_{t+1} - R_{10\text{yr},t+1}$, then this would reveal $\text{var}_t^*(R_{t+1} - R_{10\text{yr},t+1})$, and hence (in conjunction with index options and bond options) the covariance $\text{cov}_t^*(R_{t+1}, R_{10\text{yr},t+1})$.

But such markets do not currently exist, and in the meantime there is no easy solution to this problem. One pragmatic response is to assume that the relevant returns are jointly lognormal. In this case, risk-neutral and true correlation are equal to each other, as noted by Gormsen et al. (2021) (see Section 2.2.3), so if, for example, $R_{i,t+1}$ and R_{t+1} are jointly lognormal then the risk-neutral covariance that arises on implementing equation 49 with $X_{t+1} = R_{i,t+1}$ can be written as

$$\text{cov}_t^*\left(R_{t+1}^\gamma, R_{i,t+1}\right) = \rho_t \sigma_t^*(R_{t+1}^\gamma)\sigma_t^*(R_{i,t+1}). \tag{53}$$

This expresses risk-neutral covariance as the product of true correlation, which might be proxied by a backward-looking historical estimate, and two risk-neutral volatilities that are each observable (using options on the index and on asset i, respectively).

Della Corte et al. (2024) take this approach to forecast currency movements from the perspective of an investor with power utility and risk aversion $\gamma > 1$. (This exercise can be motivated by the coefficient estimates in Table 2, which are significantly larger than one in the presence of fixed effects.) As the approach is broadly applicable, requiring only that options on the appropriate asset are traded, it could for example be used to estimate inflation risk premia, bond risk premia, commodity risk premia and so on.

The lognormality assumption is a strong one, however, and one has to estimate correlations whose sizes may shift quickly, notably at times of market turmoil. Even the signs of correlations may switch: Campbell et al. (2017) show that the realized correlation between bond and stock returns was positive from around 1980 to the late 1990s, switched sign several times between 1995 and 2008, and was generally negative from 2008 to 2015.

Martin and Shi (2024) propose a different way to deal with—or rather to avoid dealing with—the correlation structure. They consider the problem of forecasting crashes in individual stocks and allow the representative agent who holds the market to have power utility. Applying equation 49 with $X_{t+1} = \mathbf{1}_{R_{i,t+1}<x}$, where the size of x indexes the severity of the crash, we have

$$\mathbb{P}_t\left(R_{i,t+1} < x\right) = \frac{\mathbb{E}_t^*\left(R_{t+1}^\gamma \mathbf{1}_{R_{i,t+1}<x}\right)}{\mathbb{E}_t^*\left(R_{t+1}^\gamma\right)}. \tag{54}$$

This generalizes the earlier equation 23 to allow for $\gamma \neq 1$ and for arbitrary returns $R_{i,t+1}$. Applied to the market itself (that is, with $R_{i,t+1} = R_{t+1}$), the risk-neutral expectations on the right-hand side of equation 54 are easily calculated from index option prices.

More generally, however, the risk-neutral expectation in the numerator is not pinned down by observable asset prices. Martin and Shi get around this problem by using the Fréchet–Hoeffding bounds to derive upper and lower bounds on the crash probability that are expressed in terms of the univariate (hence observable) risk-neutral distributions of the stock in question and the market. They argue on a priori grounds that the lower bound is likely to be closer to the truth than the upper bound, and find empirically that it is a highly statistically and economically significant forecaster of crashes.

4. CONCLUSION

Practitioners have long been interested in predictor variables based on asset prices. These risk-neutral quantities have the great advantage of being almost continuously observable, and they embody the collective views of market participants. They are used as indices of market expectations in several different settings. Forward rates (risk-neutral expected future interest rates) are used as indicators of future interest rates. Breakeven inflation (risk-neutral expected future inflation) is used as a measure of market-expected inflation. CDS rates (risk-neutral default probabilities) are used as measures of true default probabilities. Implied volatility (risk-neutral volatility) is used as a measure of true volatility.

It is important for financial economists to confront the fact that such variables—perhaps accompanied by an approximate mental adjustment to "allow for risk"—are far more widely used as a rough guide to expectations than are the predictions of the leading equilibrium models of the macro-finance literature. The appeal of risk-neutral quantities reflects the fact that they can be inferred from asset prices alone, without the need for infrequently updated macroeconomic or accounting data, or for the calibration of unobserved parameters.

The literature surveyed in this paper exploits asset prices in a similar way. But by taking the perspective of a risk-averse investor it injects a small amount of economics into the standard risk-neutral calculation. The resulting indicators account for market risk, and because they exploit risk-neutral measures of variance or covariance, we avoid the need to use realized variances or covariances as proxies for true forward-looking covariances, as in the conventional approach. The indicators point to risk premia that are volatile, skewed, and fat-tailed, spiking in times of crisis. As they are observable in real time, they provide useful information when information is most needed—during periods of market turmoil, or in the aftermath of major pieces of market-relevant news—and make risk premia more "visible" for policymakers, for academics, and for investors.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I thank Stefan Nagel, Lukas Kremens, Can Gao, John Campbell, Eben Lazarus, David Schreindorfer and Tobias Sichert for their comments. I am very grateful to Can Gao for his help generating Table 1 and to Ran Shi for his help generating Figures 1 and 2.

LITERATURE CITED

Ai, H. and Bansal, R. (2018). Risk preferences and the macroeconomic announcement premium. *Econometrica*, 86(4):1383–1430.

Aït-Sahalia, Y. (2002). Telling from discrete data whether the underlying continuous-time model is a diffusion. *Journal of Finance*, 57(5):2075–2112.

Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature*, 50(4):1007–1050.

Back, K., Crotty, K., and Kazempour, S. M. (2022). Validity, tightness, and forecasting power of risk premium bounds. *Journal of Financial Economics*, 144:732–760.

Backus, D., Chernov, M., and Martin, I. W. R. (2011). Disasters implied by equity index options. *Journal of Finance*, 66:1969–2012.

Baillie, R. T. and DeGennaro, R. P. (1990). Stock returns and volatility. *Journal of Financial and Quantitative Analysis*, 25(2):203–214.

Bansal, R., Kiku, D., Shaliastovich, I., and Yaron, A. (2014). Volatility, the macroeconomy, and asset prices. *Journal of Finance*, 69(6):2471–2511.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4):1481–1509.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121(3):823–866.

Bates, D. S. (1991). The crash of '87: Was it expected? The evidence from options markets. *Journal of Finance*, 46(3):1009–1044.

Bekaert, G., Engstrom, E., and Xing, Y. (2009). Risk, uncertainty, and asset prices. *Journal of Financial Economics*, 91:59–82.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–659.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *Journal of Finance*, 59(1):407–446.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22:4463–4492.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651.

Broadie, M., Chernov, M., and Johannes, M. (2007). Model specification and risk premia: Evidence from futures options. *Journal of Finance*, 62(3):1453–1490.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251.

Campbell, J. Y., Giglio, S., Polk, C., and Turley, R. (2018). An intertemporal CAPM with stochastic volatility. *Journal of Financial Economics*, 128(2):207–233.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.

Campbell, J. Y. and Martin, I. W. R. (2021). Sustainability in a risky world. NBER Working Paper No. 28899.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y., Sunderam, A., and Viceira, L. M. (2017). Inflation bets or deflation hedges? The changing risks of nominal bonds. *Critical Finance Review*, 6:263–301.

Carr, P. and Madan, D. (1998). Towards a theory of volatility trading. In Jarrow, R., editor, *Volatility: New Estimation Techniques for Pricing Derivatives*. Risk Books, London, UK.

Carr, P. and Wu, L. (2003). What type of process underlies options? A simple robust test. *Journal of Finance*, 58(6):2581–2610.

Chabi-Yo, F. and Loudis, J. (2020). The conditional expected market return. *Journal of Financial Economics*, 137:752–786.

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116:225–257.

Cieslak, A., Morse, A., and Vissing-Jorgensen, A. (2019). Stock returns over the FOMC cycle. *Journal of Finance*, 74(5):2201–2248.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1981). The relation between forward prices and futures prices. *Journal of Financial Economics*, 9:321–346.

Dahlquist, M. and Pénasse, J. (2022). The missing risk premium in exchange rates. *Journal of Financial Economics*, 143(2):697–715.

Della Corte, P., Gao, C., and Jeanneret, A. (2024). Survey expectations meet option prices: New insights from the FX markets. Working paper.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–25.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29.

Gandhi, M., Gormsen, N. J., and Lazarus, E. (2023). Forward return expectations. Working Paper.

Gao, C. and Martin, I. W. R. (2021). Volatility, valuation ratios, and bubbles: An empirical measure of market sentiment. *Journal of Finance*, 76(6):3211–3254.

Goetzmann, W. N., Kim, D., and Shiller, R. J. (2024). Emotions and subjective crash beliefs. NBER Working Paper No. 32589.

Gormsen, N. J. and Koijen, R. S. J. (2020). Coronavirus: Impact on stock prices and growth expectations. *Review of Asset Pricing Studies*, 10(4):574–597.

Gormsen, N. J., Koijen, R. S. J., and Martin, I. W. R. (2021). Implied dividend volatility and expected growth. *AEA Papers and Proceedings*, 111:361–365.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21:1455–1508.

Guo, H. and Whitelaw, R. F. (2006). Uncovering the risk-return relation in the stock market. *Journal of Finance*, 61(3):1433–1463.

Harvey, C. R. (2001). The specification of conditional expectations. *Journal of Empirical Finance*, 8:573–638.

Hillenbrand, S. (2024). The Fed and the secular decline in interest rates. *Review of Financial Studies*. Forthcoming.

Hsieh, P.-L., Huang, R. J., and Tzeng, L. Y. (2024). A forward-looking index of physical market variance. Working paper.

Kadan, O. and Tang, X. (2020). A bound on expected stock returns. *Review of Financial Studies*, 33(4):1565–1617.

Karatzas, I. and Shreve, S. E. (1998). *Methods of Mathematical Finance*. Springer-Verlag.

Keim, D. B. and Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17:357–390.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926.

Knox, B., Londono, J. M., Samadi, M., and Vissing-Jorgensen, A. (2024). Equity premium events. Working paper.

Knox, B. and Vissing-Jorgensen, A. (2024). A stock return decomposition using observables. Working paper.

Kremens, L. and Martin, I. W. R. (2019). The quanto theory of exchange rates. *American Economic Review*, 109(3):810–843.

Kremens, L., Martin, I. W. R., and Varela, L. (2024). Long-horizon exchange rate expectations. Working paper.

Long, J. B. (1990). The numeraire portfolio. *Journal of Financial Economics*, 26:29–69.

Lucca, D. O. and Moench, E. (2015). The pre-FOMC announcement drift. *Journal of Finance*, 70(1):329–371.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222.

Lustig, H., Roussanov, N., and Verdelhan, A. (2014). Countercyclical currency risk premia. *Journal of Financial Economics*, 111(3):527–553.

Martin, I. W. R. (2011). Simple variance swaps. NBER Working paper No. 16884.

Martin, I. W. R. (2017). What is the expected return on the market? *Quarterly Journal of Economics*, 132(1):367–433.

Martin, I. W. R. (2018). Options and the gamma knife. *Journal of Portfolio Management*, 44(6):47–55.

Martin, I. W. R. (2021). On the autocorrelation of the stock market. *Journal of Financial Econometrics*, 19(1):39–52.

Martin, I. W. R. and Papadimitriou, D. (2022). Sentiment and speculation in a market with heterogeneous beliefs. *American Economic Review*, 112(8):2465–2517.

Martin, I. W. R. and Shi, R. (2024). Forecasting crashes with a smile. Working paper.

Martin, I. W. R. and Wagner, C. (2019). What is the expected return on a stock? *Journal of Finance*, 74(4):1887–1929.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8:323–361.

Pastor, L., Sinha, M., and Swaminathan, B. (2008). Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance*, 63(6):2859–2897.

Roll, R. (1977). A critique of the asset pricing theory's tests: Part I: On past and potential testability of the theory. *Journal of Financial Economics*, 4(2):129–176.

Santa-Clara, P. and Yan, S. (2010). Crashes, volatility, and the equity premium: Lessons from S&P 500 options. *Review of Economics and Statistics*, 92(2):435–451.

Savor, P. and Wilson, M. (2013). How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis*, 48(2):343–375.

Savor, P. and Wilson, M. (2014). Asset pricing: A tale of two days. *Journal of Financial Economics*, 113(2):171–201.

Tetlock, P. C. (2023). The implied equity premium. Working paper.

Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics*, 148(2):131–148.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3):987–1035.

Weil, P. (1990). Nonexpected utility in macroeconomics. *Quarterly Journal of Economics*, 105(1):29–42.