The Implied Equity Premium

Paul C. Tetlock*
Columbia Business School

June 2023

Abstract

I propose and test a simple model of the equity premium implied by the prices of options on the stock market. The model assumes that markets for the stock index and its options are frictionless and complete to extract as much information as possible from their prices. Its forecasts of the equity premium are more accurate than those in prior work, especially when arbitrage costs are low. It offers new economic insights into why the premium varies, including why it increased for many years after the 2008 crisis. The model also provides a unified explanation of risk premiums for variance and higher-order moments of market returns.

*Comments welcome. I thank Jack McCoy, Matt Massicotte, and Neel Shah for outstanding research assistance. I appreciate helpful feedback from Geert Bekaert, Fousseni Chabi-Yo, Kent Daniel, Lars Lochstoer, Johnathan Loudis, Ian Martin, Stijn van Nieuwerburgh, and Tano Santos, as well as seminar participants at Columbia and Pomona. First draft: January 2023. All errors are my own. Author email: paul.tetlock@columbia.edu.

1 Introduction

The equity risk premium, the expected return of the stock market over a risk-free bond, is a key determinant of financial wealth that has far-reaching implications for the real economy. Accurate and timely estimates of the equity premium could reveal the origins of stock price fluctuations (Cochrane 2011), which have significant impacts on firm investment (Baker et al. (2003)), personal consumption (DiMaggio et al. (2020)), and monetary policy (Cieslak and Vissing-Jorgensen (2021)). Despite extensive research on the equity premium, economists' estimates of its value differ widely,[1] hindering intellectual and economic progress.

This paper proposes a new model of the equity premium that has compelling economic and econometric rationales, requires minimal assumptions, and makes accurate predictions of stock returns. The main assumptions are that markets for the stock index and its options are frictionless and complete,[2] reasonable approximations for S&P 500 index and option markets. These assumptions are necessary to ensure a coherent benchmark for the equity premium. If there are arbitrage opportunities, large trading costs, or incomplete markets, multiple values of the equity premium could be consistent with observed prices.

The model extracts as much information as possible from the prices of stocks and options. These prices depend not only on the equity premium, but also on the risk premiums for stock return variance, skewness, and all higher-order moments. Just as investors can buy stocks to earn the equity premium, they can create derivatives with payoffs based on return variance and higher-order moments from portfolios of options to earn the variance and higher-order premiums. Since the prices of these derivatives depend only on option prices, they reveal market or option-implied values of stock return variance, skewness, and higher-order moments. In complete and frictionless markets, one can infer the risk premiums on the stock market and these derivatives if one knows the allocations to these securities in a "growth-optimal" portfolio that maximizes an investor's expected long-run wealth.[3]

[1]In December 2021, the median one-year equity premium forecast is 6.0%, whereas the cross-sectional standard deviation of forecasts is 6.2% (Livingston Survey of Professional Forecasters).

[2]The market is complete over all possible realizations of market returns if options with a continuum of strike prices are available (Ross (1976)). Frictionless markets have no arbitrage, trading costs, and leverage and short sale constraints.

[3]In frictionless markets, the unique pricing kernel for any set of securities is the reciprocal of the gross return of the portfolio with growth-optimal weights on these securities (Long (1990)). This growth-optimal portfolio maximizes expected long-run wealth, which is equivalent to maximizing expected log utility of wealth. The first-order conditions of this maximization for each security show that the growth-optimal portfolio return is the reciprocal of the pricing kernel.

The central theoretical finding of this paper is that each of these risk premiums is a weighted sum of option-implied moments of returns, where the weights are positions in the growth-optimal portfolio. For instance, if the growth-optimal portfolio has positions of 130% in the stock market and −50% in variance-based derivatives, the equity premium is 1.3 times option-implied variance minus 0.5 times option-implied skewness. However, because the true growth-optimal weights depend on unobservable rational expectations of market returns, computing risk premiums still requires additional assumptions or data. To address this issue, a seminal paper by Martin (2017) imposes plausible constraints on investor preferences and wealth, which are equivalent to assuming growth-optimal portfolio weights of at least 100% on stocks and exactly 0% on variance and all higher-order derivatives, to obtain a useful lower bound on the equity premium.

Instead of limiting risk premiums by restricting growth-optimal weights, this paper estimates these weights using real-time data. The same growth-optimal weights that determine the equity premium also dictate the variance premium, which one can directly measure as option-implied variance minus expected market return variance. High-frequency stock prices allow for precise estimation of expected variance (Andersen et al. (2001)). Applying the main theoretical result, the expected variance premium depends linearly on observable option-implied moments of returns. I estimate this linear relationship to recover growth-optimal portfolio weights and thus market risk premiums using real-time data on stock returns and option prices. These regressions include a finite number of option-implied moments, up to four, to approximate the growth-optimal portfolio and risk premiums.

The upshot is a new framework that enables point estimates, not bounds, of option-implied risk premiums with minimal assumptions about investor preferences and wealth and minimal data requirements. These implied risk premiums are firmly grounded in asset pricing theory and forecast empirical equity and variance risk premiums quite well. Option-implied forecasts of the equity and variance risk premiums outperform those from Martin's lower bound at horizons ranging from monthly to annual.[4] The improvement in realized equity premium predictability is substantial with a quarterly out-of-sample R^2 of 3.04% for the implied equity premium as compared to 0.99% for the lower bound forecast. The model fits the expected variance risk premium even better, attaining an R^2 of 70.0% as compared to 2.5% from forecasts of the quarterly premium based on the lower bound's key assumption.

[4]I compute risk premium forecasts for the lower bound from the log utility pricing kernel implied by the exact version of Martin's lower bound, $1/R_m$, where R_m is the gross market return.

Stock return predictability from the implied equity premium is economically large, too, with monthly (annual) forecasts enabling an increase of 14% (13%) in a log utility investor's annualized expected returns. These economic magnitudes are statistically imprecise because the sample period of 1997 to 2021 is short and stock returns are highly volatile. However, consistent with theory, excluding a 61-day period in 2008 in which the model's assumption of costless arbitrage is violated significantly strengthens return predictability from the implied equity premium in statistical and economic terms.

The average implied equity premium is 7.9% per year with 4.5% volatility, whereas Martin's lower bound averages 4.2% per year with 2.2% volatility. Consistent with Martin's preference and wealth restrictions, the implied equity premium exceeds the lower bound on 99.6% of days; and it is greater than 1% below the lower bound on all days. These rare and small bound violations could arise from estimation error, approximation error, or illiquidity in the prices of long-horizon options. The largest deviations between the lower bound and implied equity premium forecasts occur in the post-2008 period, when the average deviation increases from 2.3% to 5.0% per year, which is larger than the average lower bound.

To enhance economic intuition, I interpret the growth-optimal portfolio and the implied equity premium as arising from a heterogeneous agent model like the classic behavioral models of Shiller (1984) and Campbell and Kyle (1993). The estimated growth-optimal portfolio weights are the stock and option positions chosen by an unconstrained rational log utility investor. The equity premium represents the compensation required by this growth-optimal investor for bearing stock and derivative market risk that other investors, hereafter "behavioral" investors, do not hold in equilibrium. Behavioral investors include any individuals or institutions with different preferences from log utility of wealth, biases in beliefs, or constraints on investments.

One can infer the portfolio weights chosen by behavioral investors from market clearing for stock and option markets. For example, if the growth-optimal weight on the stock market exceeds 100%, as it usually does empirically, then behavioral investors must hold less than 100% weight on stocks, consistent with their risk aversion exceeding that of log utility. If the growth-optimal weight on variance derivatives is negative, as it always is empirically, behavioral investors must hold a positive weight in these securities, consistent with a desire to hedge market variance risk. Interestingly, empirical estimates of the growth-optimal weight on skewness securities imply that behavioral investors have exposure to market crashes, much like many popular hedge fund strategies with negative skewness (Mitchell and Pulvino (2001) and Malliaris and Yan (2021)).

This heterogeneous agent perspective helps to explain why the empirical equity premium is so high on average and why it varies over time. The high average premium arises because behavioral investors are averse to stock market risk, as well as variance risk, relative to the growth-optimal investor, who must bear a disproportionate share of these risks. The premium varies over time as behavioral investors' beliefs, preferences, and constraints vary. If behavioral investors become irrationally pessimistic and demand fewer stocks, the equilibrium premium increases to entice the growth-optimal investor to buy stocks. Empirical estimates of the growth-optimal weight on stocks suggest that this mechanism can explain the persistent post-2008 increase in the equity premium. Although the model estimates these time-varying portfolio weights solely from variance premium regressions on option-implied moments, they are remarkably consistent with survey data on individual investors' beliefs and stock market participation (Greenwood and Shleifer (2014) and Gallup (2022)).

The implied risk premium model fundamentally links the risk premium in the stock market to the variance and higher-order risk premiums in the option market. Because the same time-varying growth-optimal portfolio weights (i.e., pricing kernel parameters) govern these risk premiums, they exhibit large common time-series variation. This strong link is consistent with the significant predictability of market returns from the variance premium and tail risk (Bollerslev, Tauchen, and Zhou (2009) and Bollerslev, Todorov, and Xu (2015)). The model also accounts for the average magnitude of the expected variance premium and most of its variation over time with a median R^2 of 69% across horizons from monthly to annual.[5]

Furthermore, the model correctly predicts the existence and sign of the equity risk premium's dependence on higher-order moments of stock returns. The reason is that the pricing kernel is the reciprocal of the growth-optimal portfolio return, implying that it depends nonlinearly on stock returns with alternating signs on higher-order powers of returns, which I show using a Taylor expansion in Section 2. As a result, expected stock returns depend on higher-order return moments with alternating signs, consistent with the empirical pricing of volatility, coskewness, and cokurtosis risks identified by Ang et al. (2006), Harvey and Siddique (2000), and Christoffersen et al. (2021), respectively. Because of this nonlinearity in the pricing kernel, the model also correctly predicts that the Capital Asset Pricing Model (hereafter CAPM; Sharpe (1964) and Lintner (1965)) will fail to account for empirical risk premiums since the CAPM pricing kernel is linear in market returns.

[5]For evidence that the expected variance premium is positive, see Coval and Shumway (2001), Bakshi and Kapadia (2003), Carr and Wu (2009), and Christoffersen, Heston, and Jacobs (2013).

This paper's main contribution is the implied risk premium model and estimation methodology. The model enables new estimates of all market-related risk premiums, including the equity and variance premiums, as it yields the empirical projection of the pricing kernel on to market returns. This paper differs from prior studies by exploiting the relationship between the variance risk premium and risk-neutral moments of excess market returns to obtain explicit estimates of the market pricing kernel. This kernel provides point estimates of the equity premium and insights into the pricing of market risks throughout the economy.

Recent studies by Ross (2015), Borovicka, Hansen, and Scheinkman (2016), Schneider and Trojani (2019), and Jensen, Lando, and Pedersen (2019) specify conditions that enable recovery of "physical" probabilities solely from derivative prices. The implied risk premium model proposed here augments the prices of market-related derivatives with accurate estimates of physical expected market variance. I show that this additional information enables approximate recovery without restrictive assumptions on investor preferences and wealth.

The implied risk premium model is related to recent studies that provide bounds on the equity premium, such as Martin (2017), Schneider and Trojani (2019), Chabi-Yo and Loudis (2020), and Kadan and Tang (2020). The implied risk premium approach is most similar to the Martin (2017) and Chabi-Yo and Loudis (2020) models, but those studies do not use information from the variance risk premium, which enables nearly exact recovery. This study also is related to studies by Schneider (2019) and Beason and Schreindorfer (2022) that use options to decompose the equity premium and realized market returns. I propose a novel decomposition of the equity premium in Section 4 that builds on these studies.

Many prior studies examine whether risk-neutral moments, such as skewness and kurtosis, and tail risks in options contribute to stock risk premiums. Bollerslev and Todorov (2011) relate tail risks in options to the equity premium. Kraus and Litzenberger (1978) and Harvey and Siddique (2000) show that coskewness risk has a negative price of risk in the cross-section of stocks. Neuberger (2012) and Kozhan et al. (2013) examine risk premiums related to skewness. The implied risk premium framework provides a unified explanation for many of these findings and predicts that these risk premiums should have a strong common component, consistent with evidence on the near-perfect correlation between the variance and skewness risk premiums in Kozhan et al. (2013).

2 Theory

Here I introduce the model. I derive the equilibrium pricing kernel or stochastic discount factor with minimal assumptions. In Section 2.2, I interpret this pricing kernel as arising from the interaction between rational and behavioral investors.

2.1 Pricing Securities Based on the Stock Market

The model features risky securities with returns that depend on the stock market's return. In each period t, investors can invest in an asset with a risk-free return of $R_{f,t}$, which is known at time t, and the risky market portfolio, which has an uncertain return of $R_{m,T}$ where $T > t$. The excess return of the stock market is then $\tilde{R}_{m,T} = R_{m,T} - R_{f,t}$. The model also includes markets for $K - 1 > 0$ derivative securities that offer excess returns of $\tilde{R}^k_{m,T} - c_k$, where $k = 2, 3, ..., K$ and c_k is a constant ensuring that the cost of derivative securities is zero as discussed below. I define $c_1 = 0$ so that the market's excess return has the same form as derivative excess returns, $\tilde{R}^1_{m,T} - c_1$, with $k = 1$.

These securities are like futures on the market ($k = 1$) and swaps based on market variance, skewness, and kurtosis ($k = 2, 3, 4$) and so on. The spanning result of Bakshi and Madan (2000), as specialized in Carr and Madan (2001), shows that combinations of options with a continuum of strike prices, all maturing at time T, can achieve any payoff based on the market return, $R_{m,T}$.[6] The availability of options with a continuum of strike prices completes the market, which is equivalent to assuming $K \to \infty$ securities where all higher-order moments of market excess returns are tradable. To simplify the theory, I initially assume that $K \to \infty$ to approximate the many strike prices of traded options on the stock market. For the empirical implementation, I consider approximations of this economy with finite $K \leq 4$.

I assume that there are no risk-free arbitrage opportunities and no trading frictions, such as transaction costs or restrictions on leverage or short sales. I also assume that gross market returns are finite and strictly positive, meaning at least one stock in the market retains positive value. Under these conditions, Long (1990) shows that there exists a strictly positive pricing kernel given by the reciprocal of the return on the portfolio combining the

[6] Martin (2017) applies this result to replicate the squared market return, $R^2_{m,T}$.

tradable securities with the maximum expected log return:

$$M_T = \left[R_{f,t} + \sum_{k=1}^{\infty} w_{k,t} \left(\tilde{R}_{m,T}^{k} - c_k\right)\right]^{-1}, \tag{1}$$

where $w_{k,t}$ is the weight of security k in this portfolio.[7] This portfolio is growth-optimal, hereafter GO, since it maximizes expected long-run growth of investor wealth.

Under these minimal assumptions, the GO portfolio prices all tradable securities related to the stock market:

$$\mathbb{E}_t \left[M_T R_T\right] = 1, \tag{2}$$

where R_T is the return of any of the K market-related securities or the risk-free rate. Applying this GO pricing kernel equation to the risk-free rate, I obtain the familiar identity: $\mathbb{E}_t M_T = R_{f,t}^{-1}$.[8]

One can express market prices in terms of "risk-neutral" expectations denoted by $\mathbb{E}_t^*$ and defined by the change of measure $M_T / \mathbb{E}_t M_T > 0$. The risk-neutral expectation of any tradable return is $\mathbb{E}_t^* R_T = R_{f,t} \mathbb{E}_t \left[M_T R_T\right] = R_{f,t}$. Subtracting $R_{f,t}$ from both sides, the risk-neutral expectation of any tradable excess return is zero:

$$\mathbb{E}_t^* \tilde{R}_{k,T} = 0, \forall k, \tag{3}$$

where $\tilde{R}_{k,T}$ denotes the excess return of any market-related security. Using this formula, I solve for c_k to obtain:

$$c_k = \mathbb{E}_t^* \tilde{R}_{m,T}^{k}, \forall k. \tag{4}$$

For the market, $c_1 = 0$ because $\mathbb{E}_t^* \tilde{R}_{m,T} = 0$. For the variance-based security, $c_2 = \mathbb{E}_t^* \tilde{R}_{m,T}^2$. These constants are known at time t because they are based on observable market prices—i.e., risk-neutral expectations. Empirically one can identify the constants from option prices, as I will show in equation (20).

[7] One can relax many of these conditions.

[8] I omit brackets in expectations that apply to a single term—i.e, $\mathbb{E}_t^* \left[\tilde{R}_{k,T}\right] = \mathbb{E}_t^* \tilde{R}_{k,T}$.

Now I relate rational expectations of returns to risk-neutral expectations, which are based on stock and option prices. The expected excess return of any market-related security is:

$$\begin{aligned} \mathbb{E}_t \tilde{R}_T &= \mathbb{E}_t \left[\frac{M_T}{\mathbb{E}_t M_T} \frac{\mathbb{E}_t M_T}{M_T} \tilde{R}_T \right] \\ &= R_{f,t}^{-1} \mathbb{E}_t^* \left[M_T^{-1} \tilde{R}_T \right] \\ &= R_{f,t}^{-1} \mathbb{E}_t^* \left[\left[R_{f,t} + \sum_{k=1}^{\infty} w_{k,t} \left(\tilde{R}_{m,T}^k - \mathbb{E}_t^* \tilde{R}_{m,T}^k \right) \right] \tilde{R}_T \right], \end{aligned} \tag{5}$$

where the last equality substitutes for M_T^{-1} and c_k using equations (1) and (4), respectively.

I apply equation (5) to the excess returns of the stock market and traded derivatives by setting $\tilde{R}_T = \tilde{R}_{m,T}^n$ for $n = 1, ..., K$ and simplifying:

$$\mathbb{E}_t \tilde{R}_{m,T}^n - \mathbb{E}_t^* \tilde{R}_{m,T}^n = R_{f,t}^{-1} \sum_{k=1}^{\infty} w_{k,t} \left(\mathbb{E}_t^* \tilde{R}_{m,T}^{k+n} - \mathbb{E}_t^* \tilde{R}_{m,T}^k \mathbb{E}_t^* \tilde{R}_{m,T}^n \right), \forall n. \tag{6}$$

The risk premium of any market-related security is linear in risk-neutral moments, where the term in parentheses is a demeaned risk-neutral moment. The only unknowns are the GO portfolio weights: $w_{k,t}$. However, since these weights vary over time and market returns are highly unpredictable, one cannot simply regress market securities' excess returns on risk-neutral moments to obtain accurate estimates of the weights.

The implied risk premium in (6) is simplest for the equity premium, $n = 1$:

$$\mathbb{E}_t \tilde{R}_{m,T} = R_{f,t}^{-1} \sum_{k=1}^{\infty} w_{k,t} \mathbb{E}_t^* \tilde{R}_{m,T}^{k+1}. \tag{7}$$

The exact version of Martin's (2017) expected equity premium formula corresponds to assuming that the stock market is the GO portfolio—i.e., setting $w_1 = 1$ and $w_k = 0$ for $k \geq 2$. In this special case, one obtains the exact version of Martin's (2017) equation:

$$\mathbb{E}_t R_{\tilde{m},T} = R_{f,t}^{-1} \mathbb{E}_t^* \tilde{R}_{m,T}^2, \tag{8}$$

where the right-hand side is the stock market's risk-neutral variance discounted by the risk-free rate. This equity premium formula applies only if a log utility investor chooses to hold a weight of 100% on stocks and 0% on all market derivatives, an unlikely special case.

The logic underlying equation (5) suggests a much less restrictive alternative to assuming that the stock market is the GO portfolio. If the GO investor exploits variation in the equity premium by timing the market, the market weight would not equal 1 and would not even be constant. To explore this further, I analyze an K^{th}-order approximation of the GO portfolio in which $w_{k,t} = 0$ for $k > K$, which implies that the inverse pricing kernel is a K^{th}-order polynomial:

$$M_T \approx \left[R_{f,t} + \sum_{k=1}^{K} w_{k,t} \left(\tilde{R}^k_{m,T} - \mathbb{E}^*_t \tilde{R}^k_{m,T} \right) \right]^{-1}, \tag{9}$$

where K is finite. The time-varying weights $w_{k,t}$ enable optimal market timing in K market-related securities—e.g., stocks and variance-, skewness-, and kurtosis-based derivatives when $K = 4$. With this approximation, the expected equity premium is:

$$\mathbb{E}_t \tilde{R}_{m,T} \quad = \quad R^{-1}_{f,t} \sum_{k=1}^{K} w_{k,t} \mathbb{E}^*_t \tilde{R}^{k+1}_{m,T}. \tag{10}$$

Hereafter I refer to the model of the pricing kernel in equation (9) as the implied risk premium (IRP) and its application to the stock market in equation (10) as the implied equity premium (IEP). The IRP generalization of Martin (2017) shows that the predictive power of risk-neutral variance will depend on the weight of the market, $w_{1,t}$, and any nonzero weights on market-related derivative securities in the GO portfolio. For example, a nonzero quadratic term in the pricing kernel implies that the GO portfolio has a nonzero weight in variance derivatives—i.e., $w_{2,t} \neq 0$. If all additional GO weights are zero, the expected equity premium would be linear in risk-neutral variance and skewness:

$$\mathbb{E}_t \tilde{R}_{m,T} \quad = \quad R^{-1}_{f,t} \left[w_{1,t} \mathbb{E}^*_t \tilde{R}^2_{m,T} + w_{2,t} \mathbb{E}^*_t \tilde{R}^3_{m,T} \right].$$

If instead one allows for nonzero weights on higher-order derivatives, such as bets on market skewness, the expected equity premium would be linear in additional risk-neutral moments, such as kurtosis. Since trading strategies based on skewness or even kurtosis in market returns are feasible using deep-out-of-the-money options, this study examines approximation degrees up to $K = 4$.

The remaining challenge is estimating up to $K = 4$ unknown GO weights in the IRP.[9]

[9] One could estimate predictive regressions of excess market returns on risk-neutral variance, skewness, kurtosis, and fifth moments. But the regression coefficients would be poorly estimated since realized market returns are weakly correlated with expected returns.

To tackle this problem, I exploit the pricing equation for the variance risk premium, which enables estimates of the unknown weights in terms of observable option prices and precisely estimated expected physical variance. Applying the risk premium equation (6) to market variance, $n = 2$, and setting a finite K value to approximate the GO portfolio, I obtain:

$$\mathbb{E}_t^* \tilde{R}_{m,T}^2 - \mathbb{E}_t \tilde{R}_{m,T}^2 = -R_{f,t}^{-1} \sum_{k=1}^{K} w_{k,t} \left(\mathbb{E}_t^* \tilde{R}_{m,T}^{k+2} - \mathbb{E}_t^* \tilde{R}_{m,T}^{k} \mathbb{E}_t^* \tilde{R}_{m,T}^2 \right), \tag{11}$$

an expression for the variance premium implied by option prices.

Equation (11) shows that the implied variance premium equals a linear combination of higher-order risk-neutral moments weighted by the GO portfolio weights. As discussed in Section 3, I obtain precise estimates of the left-hand side variables using realized high-frequency moments for physical variance and option prices for risk-neutral variance. I also use option prices to measure the risk-neutral moments on the right-hand side. Since equation (11) holds at all times t, I use rolling time-series regressions of the variance premium on higher-order risk-neutral moments to estimate the coefficients, $w_{k,t}$, in the GO portfolio, as discussed in Section 3.[10]

To provide intuition for the IEP, consider the linear, $K = 1$, approximation of the GO pricing kernel. Setting $K = 1$ in equation (11), the GO portfolio weight on the market is the variance risk premium divided by negative risk-neutral skewness:

$$w_{1,t} = \frac{\mathbb{E}_t^* \tilde{R}_{m,T}^2 - \mathbb{E}_t \tilde{R}_{m,T}^2}{-R_{f,t}^{-1} \mathbb{E}_t^* \tilde{R}_{m,T}^3}.$$

Since empirical estimates of the variance risk premium and negative risk-neutral skewness are both consistently positive, the estimate of the GO market weight is always positive. Using this optimal market weight, one obtain a closed-form expression for the equity premium in terms of the variance premium, risk-neutral variance, and risk-neutral skewness:

$$\mathbb{E}_t \tilde{R}_{m,T} = \frac{\mathbb{E}_t^* \tilde{R}_{m,T}^2 - \mathbb{E}_t \tilde{R}_{m,T}^2}{-\mathbb{E}_t^* \tilde{R}_{m,T}^3} \times \mathbb{E}_t^* \tilde{R}_{m,T}^2.$$

The IEP will be positive whenever the GO market weight is positive, which it is in general.

[10]Instead one could use a purely cross-sectional approach by applying equation (5) to derivatives based on market skewness, kurtosis, and higher-order moments, $R_T = \tilde{R}_{m,T}^k$, with $k \geq 3$. The empirical challenges are obtaining accurate estimates of higher-order physical moments of market returns and dealing with illiquidity in the options needed for very high-order risk-neutral moments.

In fact, as noted in the introduction, the empirical GO market weight almost always exceeds one, which is consistent with the Martin lower bound in equation (8).

For readers seeking more intuition and detail, Appendix A provides a simple two-state example that illustrates how empirical estimates of the variance premium and risk-neutral skewness enable recovery of the true equity premium and pricing kernel. The next subsection facilitates interpretation for the general case with any distribution of market returns.

2.2 A Heterogeneous Agent Model of Market-Related Securities

Building on the IRP framework, I provide a model that rationalizes the proposed GO pricing kernel and offers insights into the meaning of the parameters. This model exploits the result that a GO kernel prices a set of securities, e.g., the stock market and options on the market, if and only if it satisfies the first-order condition (FOC) from a portfolio optimization of a log utility investor who invests in these securities. Thus, the partial equilibrium model analyzes the hypothetical portfolio choices of log utility or growth-optimal (GO) investors in frictionless, complete markets for market-related securities.

There are two types of investors: type G, growth-optimal, who seek to maximize expected log utility of (long-run) wealth in period T and have rational expectations of returns; and type B who select portfolios based on possibly biased expectations of returns, non-standard preferences, and unspecified constraints. Because maximizing expected log utility is equivalent to maximizing expected long-run wealth, type G investors choose to hold the GO portfolio.[11] Although this specification of sophisticated investors has strong normative appeal (e.g., Markowitz (1976)) and support from evolutionary arguments, the existence of such investors is not necessary for the IRP model's validity. In period t, the endowed wealth of type G is $e_{G,t}$ and that of type B is $e_{B,t}$. I normalize the share of stock market shares to 1 and the supply of each derivative's shares to zero.

This interpretation of the IRP model is notable for what it does not assume. There are no restrictions on type B investors' beliefs, preferences, or constraints, and no distributional assumptions about returns, which need not even be stationary. The tractability of frictionless, complete markets and myopia of log utility investors enables this generality.

[11] Allowing for intermediate consumption has no impact on the results.

Type G (GO) investors solve the following portfolio choice optimization:

$$\max_{w_{G,k,t}} \quad \mathbb{E}_t \left[\ln \left(e_{G,t} R_{G,p,T} \right) \right], \forall k = 1, ...K, \tag{12}$$

$$\text{subject to} \quad R_{G,p,T} = R_{f,t} + \sum_{k=1}^{K} w_{G,k,t} \left(\tilde{R}_{m,T}^{k} - \mathbb{E}_t^* \tilde{R}_{m,T}^{k} \right), \tag{13}$$

where $R_{G,p,T}$ is the uncertain return on the type G (GO) portfolio. The FOC for optimal weights shows that the reciprocal of the GO portfolio return prices all assets:

$$\mathbb{E}_t \left[R_{G,p,T}^{-1} \left(\tilde{R}_{m,T}^{k} - \mathbb{E}_t^* \tilde{R}_{m,T}^{k} \right) \right] = 0, \forall k = 1, ...K, \tag{14}$$

where $k = 1$ is the FOC for the market. These FOCs imply that the risk-free rate satisfies:

$$\mathbb{E}_t R_{G,p,T}^{-1} = R_{f,t}^{-1}. \tag{15}$$

From these FOCs, the GO pricing kernel, $M_T = R_{G,p,T}^{-1}$, prices all assets because:

$$\begin{aligned} \mathbb{E}_t \left[M_T \left(\tilde{R}_{m,T}^{k} - \mathbb{E}_t^* \tilde{R}_{m,T}^{k} \right) \right] &= 0 \\ R_{f,t}^{-1} \mathbb{E}_t^* \tilde{R}_{m,T}^{k} - R_{f,t}^{-1} \mathbb{E}_t^* \tilde{R}_{m,T}^{k} &= 0, \forall k = 1, ...K, \end{aligned} \tag{16}$$

where $R_{f,t} = \left(\mathbb{E}_t M_T \right)^{-1}$.

Since the asset pricing equations ((3) and (16)) and pricing kernel equations ((1) and (13)) are the same, the solution to the log utility investor's portfolio choice problem results in the same GO pricing kernel as the no-arbitrage argument from Section 2. Therefore the same equity premium and variance risk premium formulas apply to the partial equilibrium considered here.

The market clearing equations in this heterogeneous agent model state that type G (GO) investors must hold whatever stocks and derivative securities that type B investors do not hold at equilibrium prices. These equations relate type G investors' portfolio weights, the unknown parameters in the GO pricing kernel, to interesting economic variables. Stock market clearing implies:

$$\begin{aligned} w_{G,1,t} e_{G,t} + w_{B,1,t} e_{B,t} &= P_t \\ w_{G,1,t} &= e_{G,t}^{-1} \left(P_t - w_{B,1,t} e_{B,t} \right), \end{aligned} \tag{17}$$

where P_t is the price and total capitalization of the stock market. Market clearing in each derivative market implies:

$$w_{G,k,t} \quad = \quad -e_{G,t}^{-1} w_{B,1,t} e_{B,t}, \forall k = 2, ...K. \tag{18}$$

Market clearing links GO weights to the portfolios of type B investors. For example, if type B investors hold stocks but have no net holdings of any derivatives, then $w_{G,k,t} = 0$ for $k = 2, ..., K$. In this case, the equity premium is proportional to risk-neutral variance, $\mathbb{E}_t \tilde{R}_{m,T} = R_{f,t}^{-1} w_{G,1,t} \mathbb{E}_t^* \tilde{R}_{m,T}^2$, and the GO portfolio return is linear in excess market returns, $R_{G,p,T} = R_{f,t} + w_{G,1,t} \tilde{R}_{m,T}$. If type B investors hold no stocks or the same stock weight as type G investors, then $w_{G,1,t} = 1$ and the equity premium satisfies Martin's exact formula in equation (8).

As another example, if type B investors hold stocks and variance derivatives, then type R investors are exposed to variance derivatives and any stocks not held by type B investors. The GO portfolio return is then quadratic in the excess market return, $R_{G,p,T} = R_{f,t} + w_{G,1,t} \tilde{R}_{m,T} + w_{G,2,t} \left(\tilde{R}_{m,T}^2 - \mathbb{E}_t^* \tilde{R}_{m,T}^2 \right)$, implying that the second-order, $K = 2$, approximation in the no-arbitrage model will exactly account for the equity premium as in equation (10). In this way, the complexity of type B investor portfolios determines the maximum degree of risk-neutral moments needed to account for the equity premium and the polynomial degree of the GO portfolio return.

2.3 Implications of the Model

The asset pricing implications of the model depend on the GO pricing kernel, which is the reciprocal of GO portfolio return. When behavioral (type B) investors hold no derivatives $w_{k,t} = 0$ for $k = 2, ..., K$, this model resembles the conditional CAPM in that the GO pricing kernel is a monotonic function of the conditional CAPM pricing kernel. To see how the nonlinear transformation of the CAPM pricing kernel affects equilibrium pricing, consider a Taylor expansion of this one-parameter GO pricing kernel around $\tilde{R}_{m,T} = 0$:

$$M_T = R_{f,t}^{-1} \left[1 - w_{1,t} \left(\tilde{R}_{m,T} / R_{f,t} \right) + w_{1,t}^2 \left(\tilde{R}_{m,T} / R_{f,t} \right)^2 - w_{1,t}^3 \left(\tilde{R}_{m,T} / R_{f,t} \right)^3 + ... \right], \tag{19}$$

where the ellipsis summarizes higher-order terms that follow the same sign-flipping pattern as above. The term that is linear in the market return is the same as that in the conditional CAPM. As in the conditional CAPM, the equilibrium risk-return trade-off varies over time.

In this model, variation in the weight of the market in the GO portfolio, $w_{1,t}$, determines the risk-return trade-off, including time-variation in the equity premium and the slope of the cross-sectional relation between securities' risk premiums and their market betas.

Interestingly, the same parameter, $w_{1,t}$, affects how higher-order moments are priced. The IRP model's distinct feature is the risk premium's dependence on higher-order moments. The quadratic term mimics the pricing kernel proposed by Harvey and Siddique (2000), who find that there is a significant price of coskewness risk coming from this term. The cubic term gives rise to a cokurtosis premium of the opposite sign, which is consistent with the theory and empirical findings in Christoffersen et al. (2021). Allowing for nonzero GO pricing kernel weights on market derivatives based on variance, skewness, and kurtosis, as I do in Section 3, gives rise to additional higher-order terms in the pricing kernel. Yet even the simplest version of the GO pricing kernel has the potential to explain many empirical shortcomings of the conditional CAPM.

3 Empirical Methodology

Here I use the IRP model to develop daily estimates of the equity and variance premiums for horizons of 30 days (monthly), 60 days (bimonthly), 90 days (quarterly), 180 days (semiannual), and 360 days (annual). There are three key steps in the estimation procedure:

1. estimating risk-neutral moments, including option-implied market variance
2. estimating expected (physical) market variance based on realized variance
3. regressing the variance premium on risk-neutral moments.

The key data are daily physical and risk-neutral moments on the stock market. I use the Standard and Poor's (S&P) 500 index as the proxy for the market. As in Martin (2017), I focus on the 1996 to 2021 period in which OptionMetrics data are available for options on the S&P 500 index. I use intraday data on market returns from 1994 to 2021 to measure realized market variance. I use the exchange-traded fund (ETF) with ticker SPY, SPDR S&P 500 ETF Trust, to measure intraday and daily market returns. As a historical equity premium estimate, I use the market minus risk-free (MktRf) factor from Ken French's website.

To measure risk-neutral moments of market returns, I use the prices of actively traded options on the S&P 500 index (ticker SPX) from OptionMetrics. I use only cash-settled European options with a.m. settlements that do not expire at quarter ends, following Martin

(2017) and others. Following the options literature (e.g., Bakshi, Kapadia, and Madan (2003), Carr and Madan (2005), Chang et al. (2013)), I employ thorough data cleaning procedures to reduce the impact of illiquidity on the estimated risk-neutral moments—see Appendix B.3 for option prices and Appendix B.2 for risk-neutral moments.

I measure physical moments of market returns using SPY ETF trades in the Trades and Quotes (TAQ) database. I use daily realized variance (RV) based on SPY ETF returns over 78 intraday intervals spaced equally in business (i.e., transaction) time throughout regular trading hours: 9:30am to 4:00pm Eastern time. Following the microstructure literature (e.g., Barndorff-Nielsen et al. (2008, 2009), Patton and Sheppard (2015)), I employ thorough data cleaning procedures, such as averaging subsample estimates, to reduce the impact of illiquidity on estimated physical return variance—see Appendix B.3 for details.

Some equity premium calculations require risk-free rates ranging from monthly to annual horizons and dividend yields on the market. I use constant maturity market yields of Treasury bills of 1, 3, 6, and 12 months from the Federal Reserve Economic Database (FRED) to match equity premium horizons. I use the average of the one-month and three-month rates as the 60-day rate. I use the one-month risk-free rate (Rf) from Ken French's website before July 31, 2001, when FRED data on the one-month yield become available.

Option implied volatility calculations require risk-free rates, dividend yields, and the underlying index value. I use risk-free rate data from OptionMetrics and interpolate from the nearest two dates for each option maturity. I use dividend yields from OptionMetrics for the ticker SPX. The SPX closing price from OptionMetrics is the price of the underlying SPX index.

3.1 Variance Premium Estimation

Estimation of the expected variance premium requires estimates of the risk-neutral and physical variance of market excess returns, which will enable estimation of the GO portfolio weights via equation (11) and thus the IEP.

3.1.1 Risk-Neutral Moments

I estimate risk-neutral moments of excess market returns, $\tilde{R}_{m,T}$, from the prices of call and put options on the market using the general formula of Bakshi and Madan (2000), as specialized in Carr and Madan (2001). This formula shows that risk-neutral expected excess market returns raised to the jth power are:

$$R_{f,t}^{-1}\mathbb{E}_t^* \tilde{R}_{m,T}^j = \frac{j!}{S_t^j}\left[\int_{F_{t,T}}^{\infty} (K - F_{t,T})^{j-2}\, C\,(K)\, dK + \int_0^{F_{t,T}} (K - F_{t,T})^{j-2}\, P\,(K)\, dK\right], \quad (20)$$

where $j = 2, 3, 4, 5, 6$ for relevant moments, S_t is the market index value, K is the strike price of call and put options priced at $C(K)$ and $P(K)$, respectively, and $F_{t,T} = R_{f,t}S_t$ is the futures price of the market for maturity T.[12] Denoting the risk-neutral moment of order j for maturity T at time t by $\mathrm{Mj}_{t,T}$—e.g., $\mathrm{M2}_{t,30}$ is risk-neutral market variance at the 30-day horizon on day t—I estimate $\mathrm{Mj}_{t,T}$ using the right-hand side of equation (20) multiplied by $R_{f,t}$. I refer to the right-hand side of equation (20) as a "discounted" risk-neutral moment, denoted by $\tilde{\mathrm{Mj}}_{t,T}$.

3.1.2 High-Frequency Identification of Expected Variance

The forecasts of market variance at horizons ranging from monthly to yearly are primarily based on a one-parameter model of daily realized variance (rv_t) as a fractionally integrated stochastic process. The key parameter is the order of fractional differencing, $0 \leq d \leq 0.5$. This fractionally integrated model exhibits the well-established long memory property of return variance (e.g., Andersen et al. (2001)) in which autocorrelations and impulse response function weights decay at a hyperbolic rate, much slower than the exponential decay in classic autoregressive (AR) models.

I assume the rv_t process satisfies:

$$(1 - L)^d rv_t = \epsilon_t, \quad (21)$$

$$rv_t = (1 - L)^{-d}\epsilon_t, \quad (22)$$

where ϵ_t is white noise and L is the lag operator defined by $Lx_t = x_{t-1}$ for any process x_t. Expanding equation (22) shows the implied hyperbolic impulse response weights:

$$rv_t = \epsilon_t + d\epsilon_{t-1} + \frac{|d(d-1)|}{2!}\epsilon_{t-2} + \frac{|d(d-1)(d-2)|}{3!}\epsilon_{t-3} + ..., \quad (23)$$

[12]This formula for moments of simple returns is analogous to the formula for the moments of log returns analyzed in Bakshi, Kapadia, and Madan (2003), except that I do not scale higher-order moments by variance.

I use maximum likelihood (ML) to estimate d on a recursive basis, each year expanding the estimation window by an additional year. I also demean rv_t on a recursive basis before applying ML estimation to obtain d_t.

Applying these real-time d_t estimates to the rv_t series enables real-time estimation of ϵ_t via equation (21) and thus real-time forecasts of rv_{t+h} at any horizon $h \geq 1$ via equation (22) or its equivalent equation (23). The recursive ML estimates of d_t are within 0.40 ± 0.03 in 24 of 26 years. These estimates are precise in all years with standard errors ranging from 0.01 to 0.03, with the latter applying only to the first few sample years.

I use forecasts of rv_{t+h} from this fractionally integrated model at horizons ranging from monthly ($\mathrm{RV}_{30} = \sum_{h=1}^{30} rv_{t+h}$) to annual ($\mathrm{RV}_{360} = \sum_{h=1}^{360} rv_{t+h}$) as the basis for computing the variance premium at these horizons. I adjust these raw fractionally integrated forecasts because they are based on intraday realized variance during the trading day, whereas the expected market variance in the variance premium is based on the variance of long-run (e.g., monthly or annual) market returns. The adjustment accounts for the overnight return period and illiquidity from short-run reversals in intraday, daily, and weekly returns that do not affect monthly and longer-run return variance. Although these two effects partially offset, it is unlikely that they exactly cancel.

I address both issues by applying a simple regression-based multiplier, κ_t, to fractionally integrated forecasts (e.g., $\mathbb{E}_t\mathrm{RV}_{30}$) to convert them into predictions of long-run return variance (e.g., $\mathrm{PVar}_{t,30} = \kappa_t \mathbb{E}_t \mathrm{RV}_{30}$). The multiplier, κ_t, is the coefficient from a regression of squared 30-day market returns on 30-day predicted realized variance based on a recursive expanding window. I omit the constant term so that the multiplier reflects the ratio of the average squared market return to the average cumulative realized variance. I use the lagged squared 30-day market return to maximize the correlation with the predicted 30-day realized variance, which also depends on (intraday) squared market returns over the same 30-day window as well as additional returns. Consistent with the offsetting effects, the average coefficient estimate is 1.02 with a volatility of 0.20. The coefficient estimate averages 0.90 in the pre-2008 period and 1.10 in the post-2008 period, indicating high-frequency return reversals are much more pronounced in the first half of the sample.

Table 1, Panel A summarizes these estimates of risk-neutral and predicted market variance ($M2$ and $PVar$), and the expected variance risk premium (i.e., the difference) at the monthly and annual horizons (30 and 360 days). It also reports realized monthly and annual equity premiums for comparison purposes. I annualize all quantities and multiply them by 100 to convert to annual percentages, except for the monthly equity premium which is in monthly percentages.

[Insert Table 1 here]

The bottom rows in Panel A show that the average estimated variance premium is positive at the monthly and annual horizons at 1.53% and 1.56%, respectively. The middle rows confirm that average risk-neutral monthly (annual) variance of 4.28% (4.30%) exceeds the corresponding predicted market variance of 2.76% (2.75%). The top row shows that the monthly equity premium averaged 0.75% with a volatility of 4.74% and an annualized variance of 2.70% (12×0.0474^2). This realized monthly market variance is almost identical to the average predicted monthly variance, PVar_{30}, of 2.76% shown in row four, providing evidence that the fractionally integrated model is well-calibrated. The variance premium estimates exhibit a very narrow range—the 5th to 95th percentile range is -0.01% to 5.25% (0.36% to 3.79%) for the monthly (annual) variance premium—showing that the model's variance prediction closely tracks market pricing of variance. Table 1, Panel B confirms that the correlations between model-predicted and risk-neutral variance are extremely high at 0.93 monthly and 0.91 annually, even though the fractionally integrated model does not use option prices.

I also find that unadjusted rv_{t+h} forecasts from the one-parameter (only d) fractionally integrated model perform well relative to two challenging benchmarks: forecasts from the well-known heterogeneous autoregressive (HAR) model proposed by Corsi (2009) and those from risk-neutral variance (M2). Table 2, Panel A compares the ability of the HAR, M2, and fractionally integrated (FI) models to predict realized variance RV_T, as measured by out-of-sample (OOS) R^2, at horizons ranging from monthly ($T = 30$ days) to annual ($T = 360$ days). The null model for computing OOS R^2 is average in-sample realized variance, which is 0.0260 annualized (16.1% volatility). The M2 model uses coefficients from a recursive linear regression of RV_T to make out-of-sample predictions. The fractionally integrated model beats the HAR model at all horizons, especially those longer beyond 30 days, even though it uses fewer parameters. It also beats the M2 model at all horizons, except monthly, despite one fewer parameter. The last row shows that all three models have negative point estimates of OOS R^2 at the annual horizon, mainly driven by two unanticipated crisis years (2008 and 2020) out of the 24 non-overlapping observations (NNObs) at the yearly horizon.

[Insert Table 2 here]

Table 2, Panel B shows the in-sample R^2 values resulting from adjusting each model's forecasts using linear transformations that best fit realized variance at each horizon. This panel shows that the HAR model could perform better if its coefficients were better calibrated, e.g., using shrinkage, but it would still underperform the fractionally integrated model at long horizons, which is striking given the fewer degrees of freedom in the fractionally integrated model. The fractionally integrated model exhibits an in-sample R^2 value of 10.2% at the annual horizon, which contrasts with the negative R^2 OOS values in Panel A.

Table 2, Panel C, which reports pairwise correlations between the model forecasts, provides arguably the strongest evidence for the fractionally integrated model. The simplest ex ante forecast of realized variance is a linear transformation of the market's pricing of variance (M2). At all horizons, the fractionally integrated model's prediction of unadjusted realized variance exhibits a correlation of 0.91 or 0.92 with M2. As noted earlier, the adjusted RV forecasts are even more highly correlated with risk-neutral moments than the unadjusted forecasts, suggesting that the multiplier parameter, κ_t, serves its intended purpose. Lastly, whereas the range of variance premium estimates in the fractionally integrated model is narrow, as noted in Table 1, there are several outliers in variance premium estimates from the HAR model, which occasionally predicts values below -10%.[13]

Based on these considerations, the main estimate of the expected variance premium for maturity T is:

$$\mathrm{VP}_{t,T} = \mathrm{M2}_{t,T} - \kappa_t \mathbb{E}_t \mathrm{RV}_{t,T}, \tag{24}$$

where $\mathbb{E}_t \mathrm{RV}_{t,T}$ comes from the fractionally integrated model of expected market variance. Figure 1 shows that the predicted variance from the fractionally integrated model closely tracks risk-neutral variance. Panel A (Panel B) illustrates this finding for variance at the monthly or 30-day (annual or 360-day) horizon. Panel C displays the resulting monthly and annual estimates of the expected variance premium, as computed in equation (24).

[Insert Figure 1 here]

The fractionally integrated model of expected market variance provides a "natural explanation" to the "puzzle" posed by Bollerslev et al. (2009), Bekaert and Hoerova (2014), and Cheng (2019), who find that the variance premium seems to be negative in periods of sudden market turmoil, such as September 2008, which would be inconsistent with investor aversion to variance risk. The fractionally integrated model of the variance premium does not have this puzzling feature and forecasts return variance just as well, if not better, than the models used in prior work, as shown in Table 2. Although the limited and noisy data on crises do not permit clear statistical inferences about which model is most accurate, the

[13] For comparability, I apply an analogous κ_t adjustment to the raw RV_T forecasts from the HAR model when computing the variance premium in the HAR model.

fractionally integrated model has strong theoretical appeal and is maximally parsimonious with its single parameter.

3.2 Equity Premium Estimation

Using empirical estimates of the expected variance premium from equation (24) and risk-neutral moments from equation (20), I now estimate the equity premium using the key IRP equation (10). As suggested by the equation, I use linear regressions of the expected variance premium on (up to four) discounted risk-neutral moments—M3, M4, M5, and M6 multiplied by $R_{f,T}^{-1}$—to estimate (up to four) GO portfolio weights as coefficients in equation (11).

The implementation hurdles are time-varying GO portfolio weights, heteroskedasticity in variance premiums, multicollinearity in risk-neutral moments, persistence in these measures, and approximation error in the theoretical variance premium equation (11). To estimate the time-varying GO weights, I use recursive regressions with exponentially declining weights on distant data. The half-life of the exponentially declining weights is 1000 calendar days, which matches the length of rolling windows used in the realized variance literature—e.g., Patton and Shepard (2015). Shortening the window captures more time-variation in GO weights, whereas lengthening it reduces estimation error of the weights. Other data- or theory-driven choices could improve on these weights—e.g., by incorporating insights from the dynamic beta literature as in Engle (2016).

To enhance efficiency and robustness to heteroskedasticity, I apply inverse variance (i.e., precision-based) weights to the errors in the least squares objective function, as in weighted least squares. I use $\mathrm{PVar}_{t,30}^{-1}$ as the precision weights.

I address multicollinearity in risk-neutral moments by imposing theory-driven assumptions to reduce dimensionality. As shown in Table 3, Panel A risk-neutral moments of different orders are strongly correlated within a given horizon. These strong correlations pose a challenge to accurate estimates of the GO portfolio weights, as multicollinearity among the regressors inflates standard errors. Table 3, Panel A shows correlations between M2, M3, M4, M5, and M6 at monthly and annual horizons.[14] Risk-neutral moments also are strongly correlated across horizons. Table 3, Panel B reports the correlations between discounted third- and fourth-order risk-neutral moments, M3 and M4, at monthly through annual horizons ($T = 30, 60, 90, 180, 360$ days). The correlations between the same-order moments with

[14]The table includes M2 for reference even though it is not a regressor in the variance premium regressions.

monthly to quarterly horizons range from 0.92 to 0.98. The correlations between the same-order moments with semiannual and annual horizons are also very high at 0.86 (0.97) for third-order (fourth-order) moments.

[Insert Table 3 here]

Motivated by theory, I impose a strong but reasonable assumption that the GO portfolio weights on each market-related security are equal across all horizons up to one year:

$$w_{k,t,T_1} = w_{k,t,T_2}, \forall T_1 = 30, ..., 360, T_2 = 30, ..., 360 \tag{25}$$

For example, the GO weights on the stock market are equal at the monthly and annual horizons; and the weights on variance-based derivatives are equal at these horizons. These cross-horizon restrictions facilitate identification of the GO weights because cross-moment correlation is much weaker across different horizons, as shown in Table 3, Panel A. Asset pricing theory suggests GO weights on a given market-related security are likely to be similar across horizons. In the model from Section 2.2, behavioral investor fear of stocks drives time-series variation in $w_{1,t,T}$ for all maturities T; and investor desire to hedge variance drives variation in $w_{2,t,T}$ for all T.[15] The pricing kernel from Martin's (2017) lower bound on the equity premium trivially satisfies the cross-horizon restriction in equation (25) because $w_{1,t,T} = 1$ for all T and $w_{k,t,T} = 0$ for all T when $k > 1$.

I adopt two methods to address approximation error from the truncation (finite K) of the expected variance premium equation (11) arising from the GO pricing kernel approximation in equation (9). I allow an intercept in each variance premium regression that accounts for the average value of omitted higher-order moments. I also evaluate how variance premium estimates improve as K increases, using values of $K = 1, 2, 3, 4$.

I jointly estimate variance premium regressions in equation (11) at all horizons, monthly through annual, to account for the strong correlation in the error terms and impose the cross-horizon restrictions in equation (25). I employ the feasible generalized least squares (FGLS) estimator described above, using weights inversely proportional to variance and declining exponentially with time. Because it allows for correlation across equations, this FGLS estimator is a weighted seemingly unrelated regression (SUR), as defined in Zellner (1962). I require a minimum of one year of data—i.e., 1996 data—for these recursive regressions, meaning that the first estimates become available in January of 1997. I use the estimated

[15] Although the model applies to a each maturity T, maturities are linked through the market clearing condition when there is common time-series variation in $w_{B,k,t,T}$ across T for any value of k, such as $k = 1$ (stock holdings) or $k = 2$ (variance-based derivative holdings).

time-varying GO weights from these regressions, along with risk-neutral moments, to estimate the time-varying variance and equity premiums at each horizon as in equations (11) and (10), respectively.

4 Risk Premium Estimates

Here I report the model's estimates of the variance and equity premiums and evaluate its performance against natural benchmarks. I begin by measuring the fit of the variance premium regression, R^2, across all horizons for a range of approximation degrees, $K = 1, 2, 3, 4$. As a benchmark for these four versions of the model, I consider the representative log utility model used for Martin's (2017) lower bound. In this log utility model, the variance premium is simply the negative of discounted risk-neutral skewness—i.e., equation (11) with $w_{1,t} = 1$ and $w_{k,t} = 0$ for all $k > 1$. This analysis shows whether variance premiums implied by various models can explain the empirical expected variance premium. It also enables tests of the validity of the cross-horizon restriction on GO weights, which necessarily impairs model performance at some horizons.

The variance premium tests examine the ability of models to explain the expected premium in equation (24), rather than the realized premium. Using the realized premium would reduce the power of the test, as shown in Table 2, Panel A. The precise estimate of the expected variance premium depends on the model of physical variance but alternative models provide similar estimates, as shown in Table 2, Panel C.

I report percentage R^2 values based on recursive rolling regressions for the expected variance premium as described in Section 3.2. Table 4 shows these R^2 values for monthly to annual horizons (rows) for the five models: log utility and the $K = 1, 2, 3, 4$ IRP models, which include K regressors consisting of risk-neutral moments of orders $3, ..., K + 2$. All four IRP models nest the log utility model. The null model for computing R^2 values is the historical mean of the expected variance premium, estimated using the same recursive procedure with the declining exponential weights as in the four IRP models. The recursive historical mean of the quarterly (90-day) variance premium ranges from a low of 0.95% to a high of 2.44% and averages 1.66% per year from 1997 to 2021, closely matching the average predicted values for the models. The recursive means are similar for other horizons.

[Insert Table 4 here]

The first column in Table 4 shows that the log utility model has modest predictive power for the empirical variance premium at monthly to quarterly horizons ($R^2 \approx 5\%$) but has no predictive ability at the semiannual and annual horizons ($R^2 < 0$). In contrast, the implied variance premium from all four IRP models explains most of the variation in the empirical variance premium, as by median R^2 statistics that exceed 50%. The lowest explanatory power of any model at any horizon is 43%. The median R^2 across horizons increases with the approximation degree of the model, K, suggesting that using a high-order approximation could be necessary to capture relevant variation in the GO pricing kernel. The $K = 4$ implied variance premium, which uses risk-neutral moments M3, M4, M5, and M6 to explain the variance premium, has a median R^2 value of 69% as compared to the median R^2 of 55% for the $K = 1$ model.

Table 4 provides evidence that the cross-horizon restriction, equation (25), imposed on GO weights affects its ability to explain variance premium variation at extreme horizons. Comparing the second through fourth rows to the first and last rows, one sees that the IRP model's R^2 is lowest at the shortest (30-day) and longest (360-day) horizons. Even so, the $K = 4$ model explains 62% of monthly and 56% of annual variance premium variation, indicating that the cross-horizon restriction is not overly burdensome.

Because the recursive variance premium regressions produce estimates on each date, I report estimates of the entire time-series of GO portfolio weights. Figure 2, Panel A shows the GO portfolio weights on the stock market, $w_{1,t}$ from all $K = 1, 2, 3, 4$ models. All models include the first ($k = 1$) term in the pricing kernel. Because of this term, which represents investor aversion to stock risk, the equity premium depends on the second-order ($k + 1 = 2$) risk-neutral moment (variance) and the variance premium depends on the third-order ($k + 2 = 3$) risk-neutral moment (skewness). In the heterogeneous agent model interpretation of the IRP from Section 2.2, the GO portfolio weight on the stock market reflects behavioral investors' reluctance to hold stocks, which gives rise to risk premiums that induce the GO investor to hold more stocks.

[Insert Figure 2 here]

Estimates of the GO weight on the stock market are always positive in all $K = 1, 2, 3, 4$ models. These estimates are stable except in the first sample year, 1997, when there is limited historical data. The market's GO weight in the $K = 1$ model is easiest to interpret because the GO investor holds no other market-related positions in this model. The market's GO weight in this model always exceeds one, suggesting that the typical investor is more risk averse than the log utility benchmark as discussed in Section 2.2.

I interpret the market's GO weight in the $K > 1$ models along with the GO investor's positions in other market-related securities, such as variance-based derivatives, which are correlated with stocks. Figure 2, Panel B shows the GO weights in the four market-related securities from the $K = 4$ model. This figure omits the first sample year, when estimates are noisy.

The GO weights on the stock market and the $k = 4$ security (swap on market kurtosis) are always positive, whereas the weights on the variance- and skewness-based market securities are consistently negative. Most weights fluctuate widely in the first few years of the sample and stabilize shortly after 2008, showing the impacts of estimation noise and uncertainty. There are strong time-series correlations among the four weights, stemming partly from the correlations among the underlying risk-neutral moments, which exhibit alternating signs as shown in Table 3, Panel A. As a result, weights with an odd degree $k = 1, 3$ are negatively (positively) correlated with weights of an even degree $k = 2, 4$ if these weights are of the same (opposite) signs. For example, the stock market ($k = 1$) weight is positively correlated with the variance-based ($k = 2$) weight because these weights have opposite signs—i.e., stock weight is positive and variance-based weight is negative.

By placing a positive weight on stocks, the GO investor benefits from the unconditional equity premium. With a negative weight on variance, the GO investor takes advantage of the variance premium. The GO weight on the stock market is persistently higher after the 2008 crisis in the $K \geq 3$ models. The counterpart to this increase in stock holdings is that behavioral investors must hold fewer stocks after the crisis. The consistently negative weight on skewness in the GO portfolio suggests that some behavioral investors exhibit a preference for negative skewness. Although the model does not explain why other investors are willing to speculate on stock market crashes, the reason cannot be rational log utility maximization. One possibility is that hedge funds prefer trading strategies with negative skewness because of reputation concerns, as argued in Malliaris and Yan (2021), and their assets under management have grown dramatically since 2008.

I now report estimates of the GO pricing kernel to elucidate the economic implications of these GO weights. Figure 3, Panel A shows GO pricing kernels based on different approximation degrees, $K = 1, 2, 3, 4$, using the median values of GO weight estimates over the entire sample. The excess market returns plotted on the x-axis span the 5th-to-95th percentile range of annual returns in the sample. The pricing kernel values range from 0.5 to 3. These positive values are consistent with the absence of arbitrage. All pricing kernels are monotonically decreasing over this range, consistent with investor risk aversion. The $K = 3$ pricing kernel takes on the most extreme values, ranging from 0.59 when the excess market return is 32.8% to 3.00 when the excess return is -26.4%, whereas the $K = 2$ kernel is the flattest. The $K = 4$ kernel has the highest curvature as measured by the ratio of slopes at extreme market returns, changing from a slope of -14.7 in bad times to a slope of just -0.32 in good times.

[Insert Figure 3 here]

Figure 3, Panel B shows the pricing kernel on a date with expected monthly returns exactly at the 95th percentile for the full ($K = 4$) model: May 25, 2010 in the aftermath of a sharp drop in stock prices. As expected, all pricing kernels are steeper on this date, especially those that include cubic and quartic terms. The $K = 3$ ($K = 4$) pricing kernel peaks at 3.65 (2.60) and has a low of 0.56 (0.68) in this time of turmoil. Outside the typical range of market returns, the pricing kernel approximation is not necessarily accurate. Non-monotonic kernels and negative values can arise in such cases.[16] Fortunately the IRP closed-form expressions for risk premiums do not require kernel estimates in such extreme scenarios.[17]

Table 5, Panel A summarizes the equity premium estimates from the four IRP models and the log utility model. The average annual (360-day) IEP estimates range from 6.82% to 9.44% in the IRP models; and the monthly (30-day) IEP averages range from 0.57% to 0.85% per month. These estimates are significantly higher than the log utility estimates of 4.22% per year and 0.35% per month. The 95th percentile of the IEP is 15.1% in the $K = 4$ as compared to just 7.6% in the log utility model, reflecting the higher skewness in the IEP model. The main reason for the higher average, volatility, and skewness of the IEP relative to the log utility model is that the estimated GO weight on the stock market consistently exceeds one. Because the risk-neutral variance is positive, volatile, and positively skewed, a higher GO weight on the market increases the average, volatility, and skewness of the IEP.

[Insert Table 5 here]

I analyze the forecasting performance of the IEP with the caveat that stock returns are highly unpredictable and the sample contains just 24 non-overlapping yearly periods. Table 5, Panel B shows out-of-sample R^2 values for the four IRP models and the log utility model at monthly to annual horizons. The null model for R^2 computation is a constant expected return equal to the sample average from 1926 to 1995—i.e., before the sample begins.[18] The log utility model has a positive R^2 at monthly to semiannual horizons, replicating the finding

[16]The estimated pricing kernels are positive and monotonic mainly because the positive GO stock market weight dominates the other weights at typical return realizations.

[17]For applications that require pricing kernel values in rare disaster scenarios, linear extrapolation of the GO kernel from the typical range of returns is probably sufficient.

[18]Using a recursively estimated historical mean would result in slightly worse performance for the null model because returns are modestly negatively autocorrelated in the sample.

in Martin (2017) and Chabi-Yo and Loudis (2020). The equity premium predictions from the first-order IRP model (IEP1) have lower R^2 values than the log utility model's prediction (RLUEP) at these horizons.

However, the equity premium predictions from the third- and fourth-order IRP models (IEP3 and IEP4) outperform the log utility predictions (RLUEP) at all horizons. With few exceptions, the IRP model predictions improve as the approximation degree of the GO pricing kernel increases. At all horizons, the IEP4 R^2 values are high, judged against the low standards of the return predictability literature (Goyal and Welch (2008, 2021)). The R^2 is 1.12% monthly, 3.04% quarterly, and 9.11% annually.

In economic terms, these IEP4 R^2 values translate into large improvements in the GO investor's expected return. I quantify the value of forecasting from the perspective of a GO (log utility) investor holding the market and risk-free asset using the simple one-period framework of Campbell and Thompson (2008). The ability to forecast the equity premium increases this investor's expected excess portfolio return by $\frac{R^2}{1-R^2}(1+S^2)$, where S^2 is the squared unconditional Sharpe ratio on the market and R^2 is the model's ability to predict market returns. Using the S^2 and R^2 values from Tables 1 and 5, the investor increases annualized returns by 14.1% (12.6%) with the monthly (annual) IEP4 forecasts. I interpret these large estimates of investment gains with caution given the short sample.

I now assess the calibration and incremental explanatory power of IEP forecasts using regressions to predict stock market returns. Ideally, the IEP would predict the realized equity premium with a coefficient of unity and no other variable, such as the variance premium, would forecast returns. Table 6 tests this hypothesis by regressing excess market returns on IEP4, the Martin (2017) lower bound (RLUEP), and the variance premium. The table shows results for horizons ranging from monthly to annual (30 to 360 days) in Panels A to E, where the independent and dependent variables in each regression have matching horizons. There are $h-1$ overlapping days in each daily predictability regression with a horizon of h days. Standard errors of regression coefficients are based on the truncated Hansen and Hodrick (1980) kernel with a bandwidth of $h-1$ days.

[Insert Table 6 here]

At all horizons, equity premium predictions from the main model (IEP4) compare well to those in Martin (2017) and those based on the variance premium. The univariate coefficients on IEP4 are statistically indistinguishable from unity with point estimates ranging from 0.64 to 1.12 at the monthly to annual horizons in the full-sample regressions in columns (1) to (3). Interestingly, when controlling for IEP4, all of the point estimates on risk-neutral variance,

$RLUEP$, and the variance premium, VP, are close to zero or negative, which contrasts with the (unreported) uniformly positive univariate coefficients on these predictors. Standard errors are high because of the short and volatile sample period and the strong correlations among the three predictors: IEP4, RLUEP, and VP.

Since the model in Section 2 assumes no arbitrage, the IEP should forecast market returns more accurately when this assumption is satisfied. Columns (4) to (6) in Table 6 show predictability regressions based on a sample that excludes 61 trading days from September 19 to December 15, 2008 in which arbitrage between stock and option markets is limited. I refer to the period without these 61 days as the no-arbitrage sample. The United States Securities and Exchange Commission began placing restrictions on short sales on September 17, 2008, which "dramatically increased bid-ask spreads for options" according to Battalio and Schultz (2011).[19]

Stock return predictability from IEP4 roughly triples when excluding the 1% of days (61 of 6,272) with severely limited arbitrage—i.e., R^2_{adj} in column (4) is three times its value in column (1)—in the specifications with monthly to quarterly horizons. The statistical significance of the coefficients on IEP4 is much stronger, too. The incremental R^2_{adj} from the other predictors remains low or even negative in the no-arbitrage period. These findings suggest that the IRP model performs better when arbitrage costs are low.

Figure 4 shows that the one-year (360-day) IEP from $K = 4$ is strongly correlated (0.88) with Martin's (2017) lower bound on the equity premium. It shows that IEP4 is greater than this lower bound on 99.6% of days from 1997 to 2021. The model deviation is, however, substantial in the post-2008 period, averaging 5.0% per year with 2.5% volatility. The main reason for this discrepancy is the post-2008 increase in the growth-optimal weight on stocks, as shown in Figure 2, which increases the required equity premium in equilibrium. The model estimates the time-varying market GO weight solely from rolling regressions of the variance premium on risk-neutral skewness, controlling for other moments. One possible economic driver of this increased GO weight is that behavioral investors had pessimistic expectations of returns after the 2008 crisis and were reluctant to hold stocks, as suggested by the heterogeneous agent model and survey evidence on beliefs (Greenwood and Shleifer (2014)).

[19] For each horizon, h, I measure the spread on a synthetic stock market index, created from S&P 500 options and risk-free Treasury bills maturity on day h, as a percentage of the index value. I compute the daily average synthetic spreads across all strike prices weighting options by dollar volume. I define the 61-day exclusion based on the first doubling of the weekly average synthetic spread, September 19, and the first halving of the monthly average spread, December 15, after the initial short sale ban, September 17, 2008.

[Insert Figure 4 here]

Figure 5 shows the IEP for models with approximation degrees $K = 1, 2, 3, 4$. Panel A (Panel B) shows the one-year (monthly) IEPs. Reassuringly, the correlations between the IEP estimates from different models are very high, exceeding 0.93 for all pairwise correlations between the $K = 2, 3, 4$ models at the monthly and annual horizons. There is also evidence for a common factor in expected returns across horizons. The cross-horizon correlations range from 0.85 to 0.88 between IEPs at the monthly and annual horizons for all models.

[Insert Figure 5 here]

Both panels indicate that the IEP rose and fell abruptly in the economic crises of 2008 and 2020. Recall that the IEP is a linear combination of risk-neutral moments weighted by the GO portfolio allocations. Since the $K = 1$ IEP exhibits this same pattern around crises, changes in the IEP term with risk-neutral variance must play a significant role. Since Figure 2 shows that the GO weight on the market did not change dramatically during these crises, changes in the risk-neutral variance must have driven these fluctuations in the IEP.

Figure 6 decomposes the IEP into contributions from each market exposure in the GO portfolio, as represented by each term in the weighted sum of risk-neutral moments in equation (10). The main contributor to the equity premium is the $k = 1$ term representing the GO investor's required compensation for bearing stock market risk. The $k = 1$ term, which is based on risk-neutral variance (M2), averages 7.9% per year with 5.5% volatility, accounting for the entire mean IEP and more than 100% of IEP variance. This term reaches peaks exceeding 40% per year during the 2008 and 2020 crises. The $k = 3$ term representing the GO investor's exposure to crash risk partially offsets the $k = 1$ term during crises. The reason is that the GO investor reduces exposure to market risk by placing a negative weight on $\tilde{R}_m^3$ in both crises, as shown in Figure 2, Panel B.

[Insert Figure 6 here]

Figure 7 shows the level and slope of the term structure of the implied equity premium (IEP). The level is IEP4_{360} in equation (9) based on the fourth-order ($K = 4$) approximation of the growth-optimal pricing kernel. The slope is the difference between the annualized one-year and one-month IEPs divided by the difference in these maturities: $(365/360 \times \text{IEP4}_{360} - 365/30 \times \text{IEP4}_{30})\,/((360 - 30)/365)$.

[Insert Figure 7 here]

The average slope of the IEP term structure is zero, reflecting an average of a slight positive slope in normal times and an extremely negative slope in times of crisis. The level and slope of the IEP term structure are strongly negatively correlated. The slope has a correlation of -0.57 with IEP4$_{360}$ and -0.93 with IEP4$_{30}$. Whenever the IEP is high, the short-run IEP is even higher than the long-run IEP. The short-run IEP exhibits the highest volatility among IEPs at different horizons—e.g., the volatilities of annualized IEP4s are $10.83\% = 365/30 \times 0.89\%$ monthly versus $4.56\% = 365/360 \times 4.50\%$ annual from Table 5, Panel A. This pattern of higher volatility for short-term rates is reminiscent of the analogous finding for bonds, but there is no notable correlation (-0.06) between the slope of the IEP and US Treasury term structures at the one-month to one-year horizon.

5 Conclusion

The economic insight from the IRP model is that the equilibrium equity premium depends on the amount of stock market risk borne by a GO investor, who is exposed to the market through stocks and market-related derivatives. If this GO investor has a real-world counterpart, one can interpret GO exposures as market positions that other investors do not want to hold. Variation in investor desires to hold market positions drives variation in the equity premium and other risk premiums related to the market, such as the variance and skewness risk premiums.

The econometric innovation in the IRP model is using high-frequency data to identify the stock market's physical return variance, which facilitates the recovery of market risk premiums from option prices. Although the market variance premium is well estimated, approximation and estimation error affect risk premium estimates through their impact on the weights in the growth-optimal pricing kernel. Future research should seek to minimize these errors by improving risk-neutral moment estimation and investigating higher-order physical moments of returns. Even without these improvements, this paper's implementation of the IRP model provides useful new estimates of market risk premiums and the empirical pricing kernel. Researchers can use these estimates to test whether market-related risks are priced in other securities, including individual stocks and bonds. Lastly, since the theory and estimation approach here applies equally well to the set of individual securities with traded options, one can use the IRP model to estimate expected returns of individual stocks with options, even without the preference assumptions in Martin and Wagner (2019) or Kadan and Tang (2020). I am pursuing this idea in ongoing research.

References

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens, 2001, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43–76.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Baker, Malcolm, Jeremy C. Stein, and Jeffrey Wurgler, 2003, When does the market matter? Stock prices and the investment of equity-dependent firms, *Quarterly Journal of Economics* 118, 969–1005.

Bakshi, Gurdip, and Nikunj Kapadia, 2003, Delta-hedged gains and the negative market volatility risk premium, *Review of Financial Studies* 16, 527—566.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics skew laws and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Bakshi, Gurdip, and Dilip Madan, 2000, Spanning and derivative-security evaluation, *Journal of Financial Economics* 55, 205–238.

Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard, 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard, 2009, Realized kernels to practice: Trades and quotes, *Econometrics Journal* 12, 1–32.

Battalio, Robert, and Paul Schultz, 2011, Regulatory uncertainty and market liquidity: The 2008 short sale ban's impact on equity option markets, *Journal of Finance* 66, 2013–2053.

Bekaert, Geert, and Marie Hoerova, 2014, The VIX, the variance premium, and stock market volatility, *Journal of Econometrics* 183, 181–192.

Beason, Tyler, and David Schreindorfer, 2022, Dissecting the equity premium, *Journal of Political Economy* 130, 2203–2222.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

Bollerslev, Tim, and Viktor Todorov, 2011, Tails, fears, and risk premia, *Journal of Finance* 66, 2165–2211.

Bollerslev, Tim, Viktor Todorov, and Lai Xu, 2015, Tail risk premia and return predictability, *Journal of Financial Economics* 118, 113—134.

Borovicka, Jaroslav, Lars P. Hansen, and Jose A. Scheinkman, 2016, Misspecified recovery, *Journal of Finance* 2493–2544.

Campbell, John Y., and Albert S. Kyle, 1993, Smart money, noise trading and stock price behaviour, *Review of Economic Studies* 60, 1–34.

Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661—76.

Campbell, John Y. and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509–1531.

Carr, Peter, and Dilip Madan, 2001, Optimal positioning in derivative securities, *Quantitative Finance* 1, 19–37.

Carr, Peter, and Dilip Madan, 2005, A note on sufficient conditions for no arbitrage, *Finance Research Letters* 2, 125—130.

Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 1311–1341.

Chabi-Yo, Fousseni, and Johnathan Loudis, 2020, The conditional expected market return *Journal of Financial Economics* 137, 752–786.

Chang, Bo-Young, Peter Christoffersen, and Kris Jacobs, 2013, Market skewness risk and the cross section of stock returns, *Journal of Financial Economics* 107, 46–68.

Cheng, Ing-Haw, 2019, The VIX premium, *Review of Financial Studies* 32, 180–227.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2013, Capturing option anomalies with a variance-dependent pricing kernel, *Review of Financial Studies* 1962–2006.

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2018, Illiquidity premia in the equity options market, *Review of Financial Studies* 31, 811–851.

Christoffersen, Peter, Mathieu Fournier, Kris Jacobs, and Mehdi Karoui, 2021, Option-based estimation of the price of coskewness and cokurtosis risk, *Journal of Financial and Quantitative Analysis* 56, 65–91.

Cieslak, Anna, and Annette Vissing-Jorgensen, 2021, The economics of the Fed put, *Review of Financial Studies* 34, 4045—4089.

Claus, James, and Jacob Thomas, 2001, Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets, *Journal of Finance* 56, 1629–1666.

Cochrane, John, 2011, Discount rates, *Journal of Finance* 66, 1047–1108.

Corsi, Fulvio, 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

Coval, Joshua D., and Tyler Shumway, 2001, Expected option returns, *Journal of Finance* 56, 983–1009.

Di Maggio, Marco, Amir Kermani, and Kaveh Majlesi, 2020, Stock market returns and consumption, *Journal of Finance* 75, 2851–3372.

Engle, Robert F., 2016, Dynamic conditional beta, *Journal of Financial Econometrics* 14, 643–667.

Fama, Eugene F., and Kenneth R. French, 2002, The equity premium, *Journal of Finance* 57, 637–659.

Gallup Organization, 2022, Washington, D.C., website retrieved March 22, 2023, available at https://news.gallup.com/poll/266807/percentage-americans-owns-stock.aspx

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

Goyal, Amit, Ivo Welch, and Athanasse Zafirov, 2021, A comprehensive look at the empirical performance of equity premium prediction II, Swiss Finance Research Institute working paper, available at SSRN: https://ssrn.com/abstract=3929119.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Hansen, Lars P., and Robert J. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Jensen, Christian S., David Lando, and Lasse H. Pedersen, 2019, Generalized recovery, *Journal of Financial Economics* 133, 154—174.

Kadan, Ohad, and Xiaoxiao Tang, 2020, A bound on expected stock returns, *Review of Financial Studies* 33, 1565—1617

Kozhan, Roman, Anthony Neuberger, and Paul Schneider, 2013, The skew risk premium in the equity index market, *Review of Financial Studies* 26, 2174—2203.

Kraus, Alan, and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *Journal of Finance* 31, 1085–1100.

Lintner, John, 1965, Security prices, risk and maximal gains from diversification, *Journal of Finance* 20, 587—615.

Long, John B., 1990, The numeraire portfolio, *Journal of Financial Economics* 26, 29–69.

Malliaris, Steven, and Hongjun Yan, 2021, Reputation concerns and slow-moving capital, *Review of Asset Pricing Studies*, 11 580–609.

Markowitz, Harry M., 1976, Investment for the long run: New evidence for an old rule, *Journal of Finance* 31, 1273–1286.

Martin, Ian, 2017, What is the expected return on the market? *Quarterly Journal of Economics* 132, 367—433.

Martin, Ian, and Christian Wagner, 2019, What is the expected return on a stock? *Journal of Finance* 74, 1887—1929.

Mitchell, Mark, and Todd Pulvino, 2001, Characteristics of risk and return in risk arbitrage, *Journal of Finance* 56, 2135–2175.

Neuberger, Anthony, 2012, Realized skewness, *Review of Financial Studies* 25, 3423–3455.

Patton, Andrew J., and Kevin Sheppard, 2015, Good volatility, bad volatility: Signed jumps and the persistence of volatility, *Review of Economics and Statistics* 97, 683–697.

Ross, Stephen A., 1976, Options and efficiency, *Quarterly Journal of Economics* 90, 75–89.

Ross, Steve, 2015, The recovery theorem, *Journal of Finance* 70, 615–648.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Schneider, Paul, 2019, Anatomy of the market return, *Journal of Financial Economics* 132, 325–350.

Schneider, Paul, and Fabio Trojani, 2019, (Almost) model-free recovery, *Journal of Finance* 74, 323–370.

Shiller, Robert J., 1984, Stock prices and social dynamics *Brookings Papers on Economic Activity* 2, 457–498.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* 57, 348–368.

Table 1: Summary Statistics

Panel A summarizes estimates of risk-neutral and predicted market variance ($M2$ and $PVar$), and the variance risk premium (i.e., the difference), and the equity premiums (EP) at the monthly and annual horizons (30 and 360 days). I annualize all quantities and multiply them by 100 to convert to annual percentages, except for the monthly EP which is in monthly percentage terms. The statistics in columns are the mean, standard deviation (StdDev), percentiles (P5, P25, P50, P75, P95), and number of days in the 1997 to 2021 sample.

	Mean	StdDev	P5	P25	P50	P75	P95	Days
EP_{30}	0.75	4.74	-7.08	-1.58	1.27	3.49	7.16	6524
EP_{360}	8.64	17.72	-26.10	2.10	11.26	18.89	33.96	6295
$M2_{30}$	4.28	4.37	1.17	1.87	3.14	5.03	10.98	6545
$PVar_{30}$	2.76	2.60	0.99	1.42	2.06	3.30	5.91	6545
$M2_{360}$	4.30	2.23	2.05	2.68	3.84	5.33	7.87	6545
$PVar_{360}$	2.75	1.41	1.41	1.83	2.38	3.21	4.90	6545
VP_{30}	1.53	2.18	-0.01	0.34	0.87	1.85	5.25	6545
VP_{360}	1.56	1.12	0.36	0.78	1.22	2.06	3.79	6545

Panel B reports correlations among estimates of risk-neutral and predicted market variance ($M2$ and $PVar$), and the variance risk premium (i.e., the difference) at the monthly and annual horizons (30 and 360 days).

	EP_{30}	EP_{360}	$M2_{30}$	$PVar_{30}$	$M2_{360}$	$PVar_{360}$	VP_{30}
EP_{30}	1.00	0.34	0.11	0.10	0.10	0.11	0.10
EP_{360}	0.34	1.00	0.22	0.26	0.18	0.29	0.12
$M2_{30}$	0.11	0.22	1.00	0.93	0.85	0.82	0.90
$PVar_{30}$	0.10	0.26	0.93	1.00	0.85	0.91	0.67
$M2_{360}$	0.10	0.18	0.85	0.85	1.00	0.91	0.69
$PVar_{360}$	0.11	0.29	0.82	0.91	0.91	1.00	0.55
VP_{30}	0.10	0.12	0.90	0.67	0.69	0.55	1.00
VP_{360}	0.06	-0.00	0.66	0.55	0.85	0.55	0.67

Table 2: Predicting Realized Variance

This table compares realized variance forecasts from the heterogeneous autoregressive (HAR), risk-neutral variance (M2), and fractionally integrated (FI) models. Panel A shows out-of-sample (OOS) R^2, at horizons ranging from monthly ($T = 30$ days) to annual ($T = 360$ days). The OOS R^2 benchmark is average in-sample realized variance, which is 0.0260 annualized (16.1% volatility). The last column shows the number of non-overlapping observations (NNObs) at each horizon in the 1997 through 2021 sample.

Horizon	HAR	M2	FI	NNObs
30	0.397	0.414	0.410	303
60	0.238	0.243	0.290	151
90	0.140	0.167	0.217	100
180	0.020	0.062	0.093	50
360	-0.083	-0.118	-0.050	24

Panel B shows the in-sample R^2 values resulting from adjusting each model's forecasts using linear transformations that best fit realized variance at each horizon.

Horizon	HAR	M2	FI	NNObs
30	0.439	0.425	0.421	303
60	0.325	0.272	0.309	151
90	0.255	0.202	0.251	100
180	0.154	0.124	0.169	50
360	0.061	0.065	0.102	24

Panel C reports pairwise correlations between the models' forecasts at each horizon.

Horizon	HAR-M2	HAR-FI	FI-M2	NNObs
30	0.860	0.904	0.917	303
60	0.810	0.863	0.916	151
90	0.780	0.834	0.918	100
180	0.714	0.785	0.916	50
360	0.620	0.759	0.906	24

Table 3: Correlations among Risk-Neutral Moments

Panel A shows correlations between risk-neutral moments of different orders—M2 (variance), M3 (skewness), M4 (kurtosis), M5, and M6—at monthly (30-day) and annual (360-day) horizons.

	$M2_{30}$	$M3_{30}$	$M4_{30}$	$M5_{30}$	$M6_{30}$	$M2_{360}$	$M3_{360}$	$M4_{360}$	$M5_{360}$
$M2_{30}$	1.00	-0.93	0.91	-0.84	0.81	0.85	-0.44	0.83	-0.46
$M3_{30}$	-0.93	1.00	-0.92	0.93	-0.86	-0.77	0.53	-0.78	0.55
$M4_{30}$	0.91	-0.92	1.00	-0.95	0.97	0.70	-0.37	0.77	-0.43
$M5_{30}$	-0.84	0.93	-0.95	1.00	-0.96	-0.65	0.42	-0.72	0.48
$M6_{30}$	0.81	-0.86	0.97	-0.96	1.00	0.59	-0.33	0.69	-0.41
$M2_{360}$	0.85	-0.77	0.70	-0.65	0.59	1.00	-0.67	0.95	-0.67
$M3_{360}$	-0.44	0.53	-0.37	0.42	-0.33	-0.67	1.00	-0.65	0.96
$M4_{360}$	0.83	-0.78	0.77	-0.72	0.69	0.95	-0.65	1.00	-0.70
$M5_{360}$	-0.46	0.55	-0.43	0.48	-0.41	-0.67	0.96	-0.70	1.00
$M6_{360}$	0.75	-0.71	0.78	-0.71	0.73	0.83	-0.55	0.96	-0.64

Panel B reports the correlations between third- and fourth-order risk-neutral moments, M3 (skewness) and M4 (kurtosis), at monthly through annual horizons (30, 60, 90, 180, 360 days).

	$M3_{30}$	$M3_{60}$	$M3_{90}$	$M3_{180}$	$M3_{360}$	$M4_{30}$	$M4_{60}$	$M4_{90}$	$M4_{180}$
$M3_{30}$	1.00	0.96	0.92	0.81	0.53	-0.92	-0.90	-0.89	-0.86
$M3_{60}$	0.96	1.00	0.98	0.89	0.63	-0.87	-0.92	-0.92	-0.91
$M3_{90}$	0.92	0.98	1.00	0.94	0.71	-0.82	-0.88	-0.91	-0.92
$M3_{180}$	0.81	0.89	0.94	1.00	0.86	-0.67	-0.75	-0.79	-0.85
$M3_{360}$	0.53	0.63	0.71	0.86	1.00	-0.37	-0.44	-0.50	-0.60
$M4_{30}$	-0.92	-0.87	-0.82	-0.67	-0.37	1.00	0.97	0.94	0.87
$M4_{60}$	-0.90	-0.92	-0.88	-0.75	-0.44	0.97	1.00	0.98	0.94
$M4_{90}$	-0.89	-0.92	-0.91	-0.79	-0.50	0.94	0.98	1.00	0.98
$M4_{180}$	-0.86	-0.91	-0.92	-0.85	-0.60	0.87	0.94	0.98	1.00
$M4_{360}$	-0.78	-0.85	-0.87	-0.82	-0.65	0.77	0.86	0.91	0.97

Table 4: Predicting the Expected Variance Premium

This table compares the fit of five models in predicting the expected variance premiums in equation (24). The fit measure is model R^2 based on recursive rolling regressions. The table shows percentage R^2 values for monthly to annual horizons (rows) for the five models (columns): representative log utility (RLUVP) and the $K = 1, 2, 3, 4$ implied variance premium (IVP) models, which include K regressors consisting of risk-neutral moments of orders $3, ..., K + 2$. The null model for computing R^2 values is the historical mean of the variance premium, using the same recency weights as in the IVP models. The last column reports the number of days in the model evaluation period spanning 1997 to 2021.

	RLUVP	IVP1	IVP2	IVP3	IVP4	Days
30	5.4	47.0	44.5	61.4	61.6	6293
60	5.9	56.1	53.4	67.9	69.4	6293
90	2.5	56.3	54.6	67.5	70.0	6293
180	-8.4	55.2	58.8	62.7	69.9	6293
360	-27.0	42.8	65.1	46.3	55.8	6293

Table 5: Summary of Equity Premium Predictions

Panel A summarizes the equity premium estimates from the four implied equity premium (IEP) models and from the representative log utility model (RLUEP) at monthly (30-day) and annual (360-day) horizons. The IEP that uses K terms to approximate the pricing kernel at horizon h days is $IEPK_h$. The RLUEP at horizon h days is $RLUEP_h$. Equity premiums are not annualized. The columns show the mean (Mean), standard deviation (StdDev), and percentiles (P5, P25, P50, P75, P95) of equity premiums for all 6,293 days from 1997 to 2021.

	Mean	StdDev	P5	P25	P50	P75	P95	Days
$RLUEP_{30}$	0.35	0.36	0.09	0.15	0.26	0.41	0.91	6293
$IEP1_{30}$	0.70	0.85	0.18	0.28	0.46	0.81	1.81	6293
$IEP2_{30}$	0.57	0.74	0.16	0.23	0.37	0.65	1.45	6293
$IEP3_{30}$	0.85	1.07	0.24	0.37	0.55	0.91	2.12	6293
$IEP4_{30}$	0.70	0.89	0.17	0.30	0.46	0.77	1.78	6293
$RLUEP_{360}$	4.22	2.16	1.95	2.69	3.83	5.12	7.63	6293
$IEP1_{360}$	8.17	5.02	3.89	4.99	6.68	10.07	15.43	6293
$IEP2_{360}$	6.82	4.33	3.43	4.17	5.35	8.29	13.10	6293
$IEP3_{360}$	9.44	5.56	4.57	6.23	7.78	10.72	18.36	6293
$IEP4_{360}$	7.88	4.50	3.42	5.19	6.74	9.38	15.10	6293

Panel B reports out-of-sample R^2 values of equity premium predictions from the IEP and RLUEP models at horizons ranging from monthly (30 days) to annual (360 days). The last column shows the number of non-overlapping observations (NNObs) at each horizon in the 1997 through 2021 sample.

Horizon	RLU	IEP1	IEP2	IEP3	IEP4	NNObs
30	0.58	-1.00	-0.22	-0.12	1.12	303
60	1.00	-1.34	-0.05	0.43	2.23	151
90	0.99	-2.25	-0.55	0.67	3.04	100
180	1.81	-0.29	1.99	6.98	8.71	50
360	-1.29	-2.47	0.61	8.71	9.11	24

Table 6: Market Return Predictability Regressions

This table shows regressions of the realized market equity premium, EP_h, the implied equity premium, $IEP4_h$, the representative log utility equity premium, $RLUEP_h$, and the variance premium, VP_h, where h is the horizon in days and the regressors are lagged by h days. Panels A to E report regressions for horizons $h = 30, 60, 90, 180, 360$ days. Hansen-Hodrick (1980) standard errors based on a bandwidth of $h - 1$ days appear in parentheses. The "Observations" row shows the number of days included in the regression. The "All" sample includes all days from 1997 through 2021, whereas the "No Arb" sample excludes 61 days from September 19, 2008 to December 15, 2008.

Panel A: 30-Day Horizon

	(1)	(2)	(3)	(4)	(5)	(6)
	EP_{30}	EP_{30}	EP_{30}	EP_{30}	EP_{30}	EP_{30}
$IEP4_{30}$	0.64	0.61	0.57	1.33***	0.89	1.28
	(0.51)	(1.20)	(0.77)	(0.48)	(1.29)	(0.90)
$RLUEP_{30}$		0.08			1.15	
		(2.91)			(2.82)	
VP_{30}			0.03			0.02
			(0.24)			(0.26)
Observations	6,272	6,272	6,272	6,211	6,211	6,211
Sample	All	All	All	No Arb	No Arb	No Arb
R^2_{adj}	0.014	0.014	0.014	0.039	0.039	0.039

Panel B: 60-Day Horizon

	(1)	(2)	(3)	(4)	(5)	(6)
	EP_{60}	EP_{60}	EP_{60}	EP_{60}	EP_{60}	EP_{60}
$IEP4_{60}$	0.68	0.58	0.69	1.31***	1.11	1.44*
	(0.61)	(1.26)	(0.92)	(0.49)	(1.21)	(0.87)
$RLUEP_{60}$		0.25			0.51	
		(2.50)			(2.50)	
VP_{60}			-0.01			-0.12
			(0.47)			(0.47)
Observations	6,251	6,251	6,251	6,190	6,190	6,190
Sample	All	All	All	No Arb	No Arb	No Arb
R^2_{adj}	0.025	0.025	0.024	0.063	0.063	0.064

Table 6 (continued): Market Return Predictability Regressions

Panel C: 90-Day Horizon

	(1) EP_{90}	(2) EP_{90}	(3) EP_{90}	(4) EP_{90}	(5) EP_{90}	(6) EP_{90}
$IEP4_{90}$	0.70	0.69	0.71	1.46***	1.38	1.71***
	(0.71)	(1.26)	(0.98)	(0.35)	(1.06)	(0.60)
$RLUEP_{90}$		0.02			0.22	
		(2.36)			(2.38)	
VP_{90}			-0.02			-0.35
			(0.67)			(0.62)
Observations	6,230	6,230	6,230	6,169	6,169	6,169
Sample	All	All	All	No Arb	No Arb	No Arb
R^2_{adj}	0.032	0.032	0.032	0.101	0.101	0.103

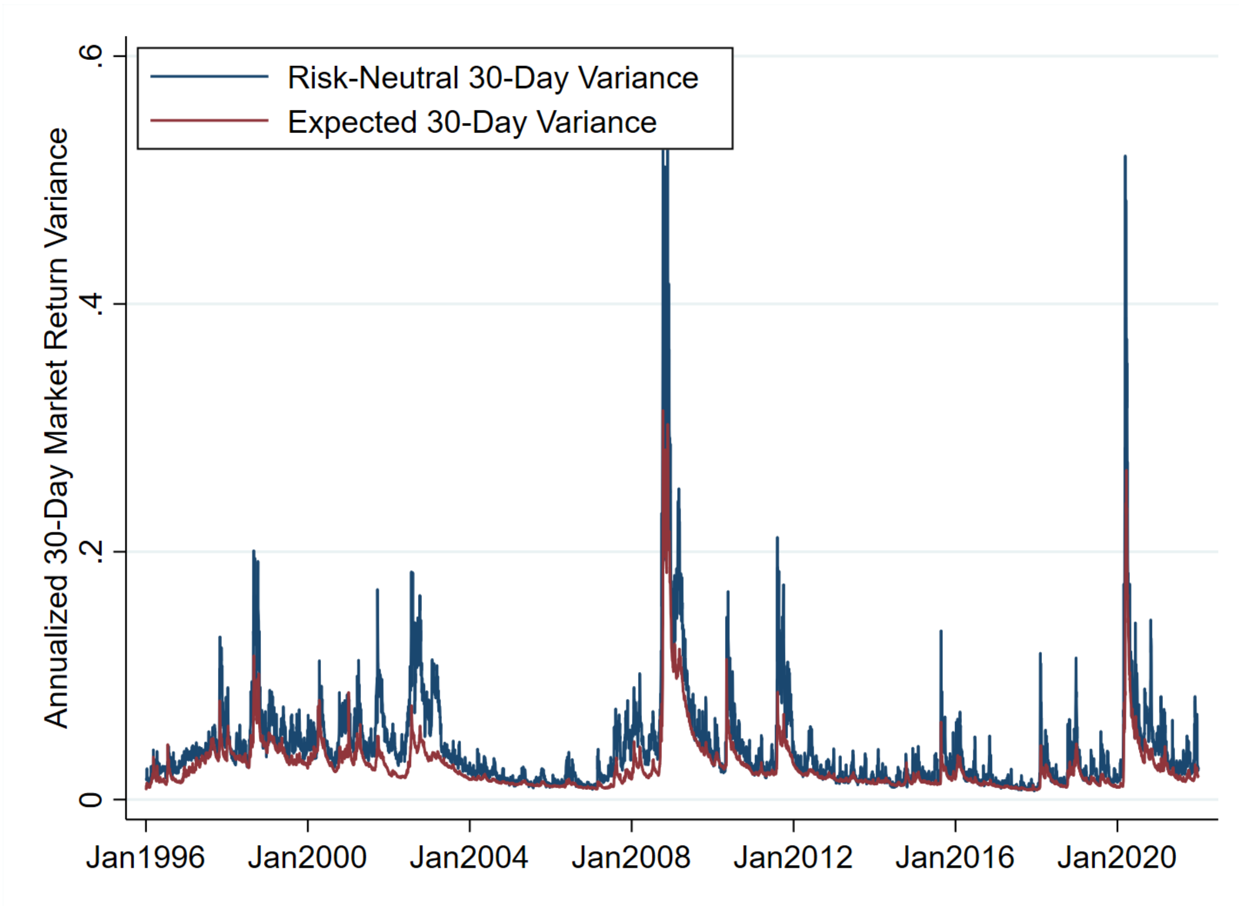
Panel D: 180-Day Horizon

	(1) EP_{180}	(2) EP_{180}	(3) EP_{180}	(4) EP_{180}	(5) EP_{180}	(6) EP_{180}
$IEP4_{180}$	1.03**	1.24	1.43**	1.46***	1.54	2.00***
	(0.49)	(1.44)	(0.58)	(0.42)	(1.39)	(0.48)
$RLUEP_{180}$		-0.52			-0.20	
		(2.82)			(2.85)	
VP_{180}			-1.18			-1.53
			(1.26)			(1.33)
Observations	6,167	6,167	6,167	6,106	6,106	6,106
Sample	All	All	All	No Arb	No Arb	No Arb
R^2_{adj}	0.080	0.080	0.089	0.122	0.122	0.136

Panel E: 360-Day Horizon

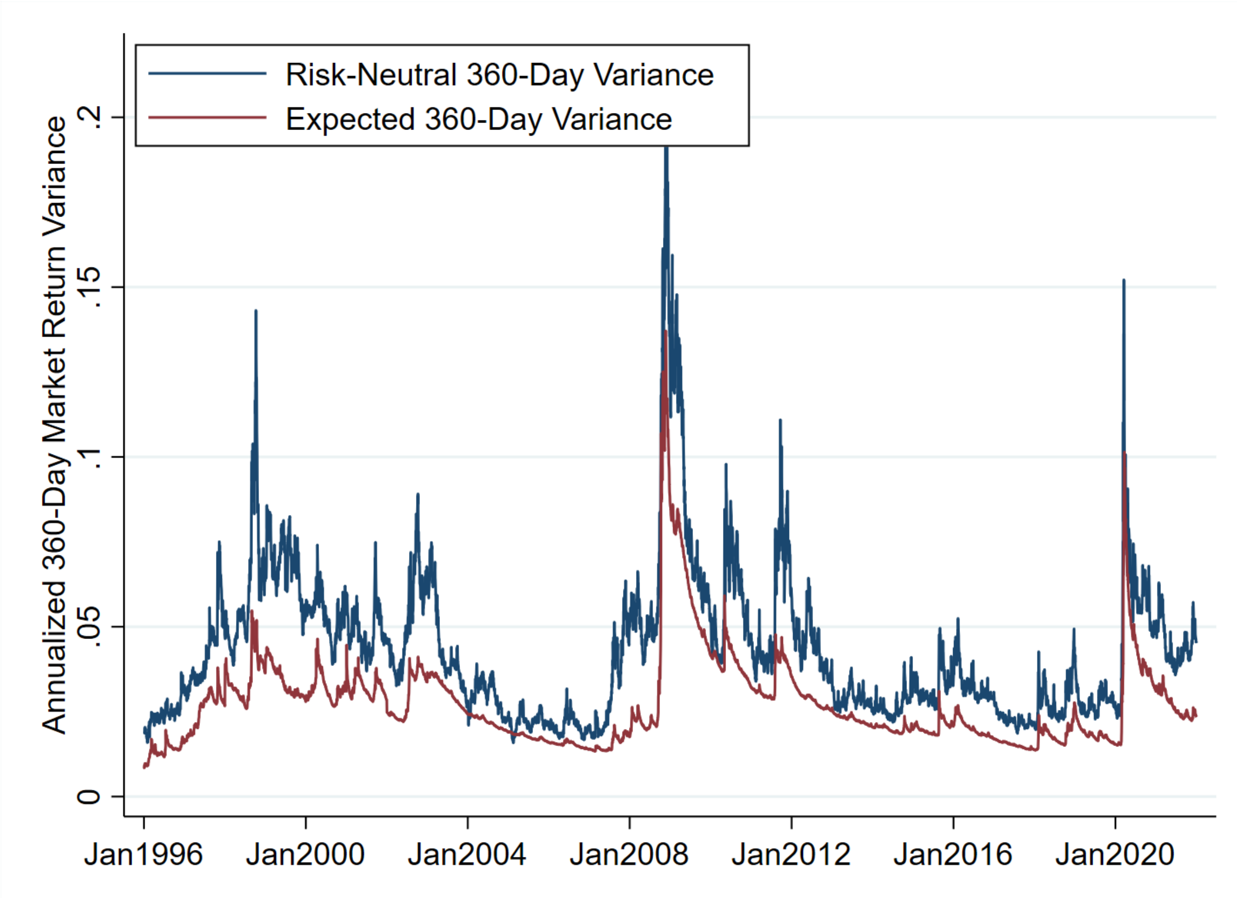
	(1) EP_{360}	(2) EP_{360}	(3) EP_{360}	(4) EP_{360}	(5) EP_{360}	(6) EP_{360}
$IEP4_{360}$	1.12**	1.40	1.97***	1.23**	1.47	2.14***
	(0.50)	(1.87)	(0.67)	(0.57)	(1.87)	(0.71)
$RLUEP_{360}$		-0.64			-0.58	
		(3.46)			(3.49)	
VP_{360}			-5.00			-5.27
			(3.34)			(3.54)
Observations	6,043	6,043	6,043	5,982	5,982	5,982
Sample	All	All	All	No Arb	No Arb	No Arb
R^2_{adj}	0.084	0.085	0.143	0.082	0.083	0.146

Figure 1: The Estimated Variance Premium



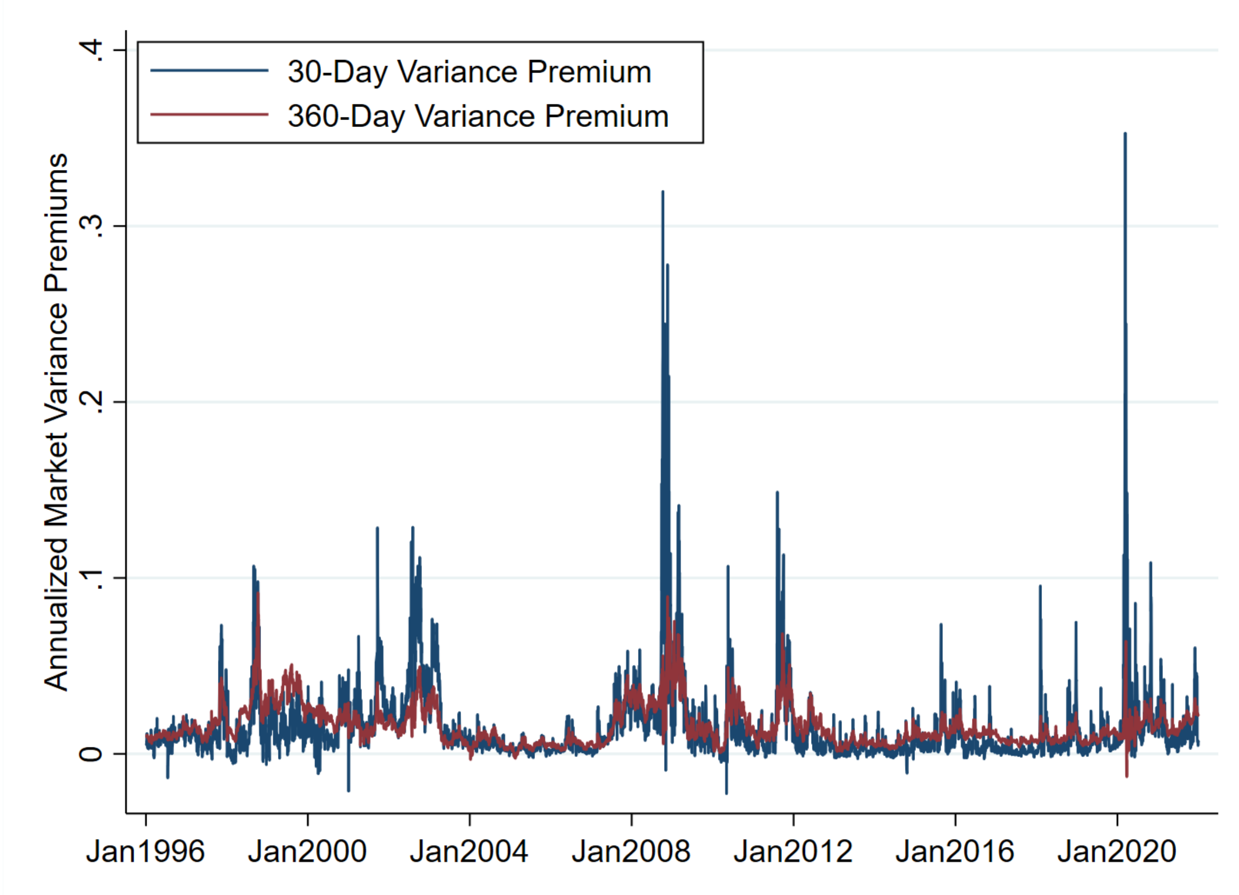
Panel A shows predicted variance from the fractionally integrated model and risk-neutral variance from option prices at the monthly (30-day) horizon. The data are daily from 1996 through 2021.

Figure 1 (continued): The Estimated Variance Premium



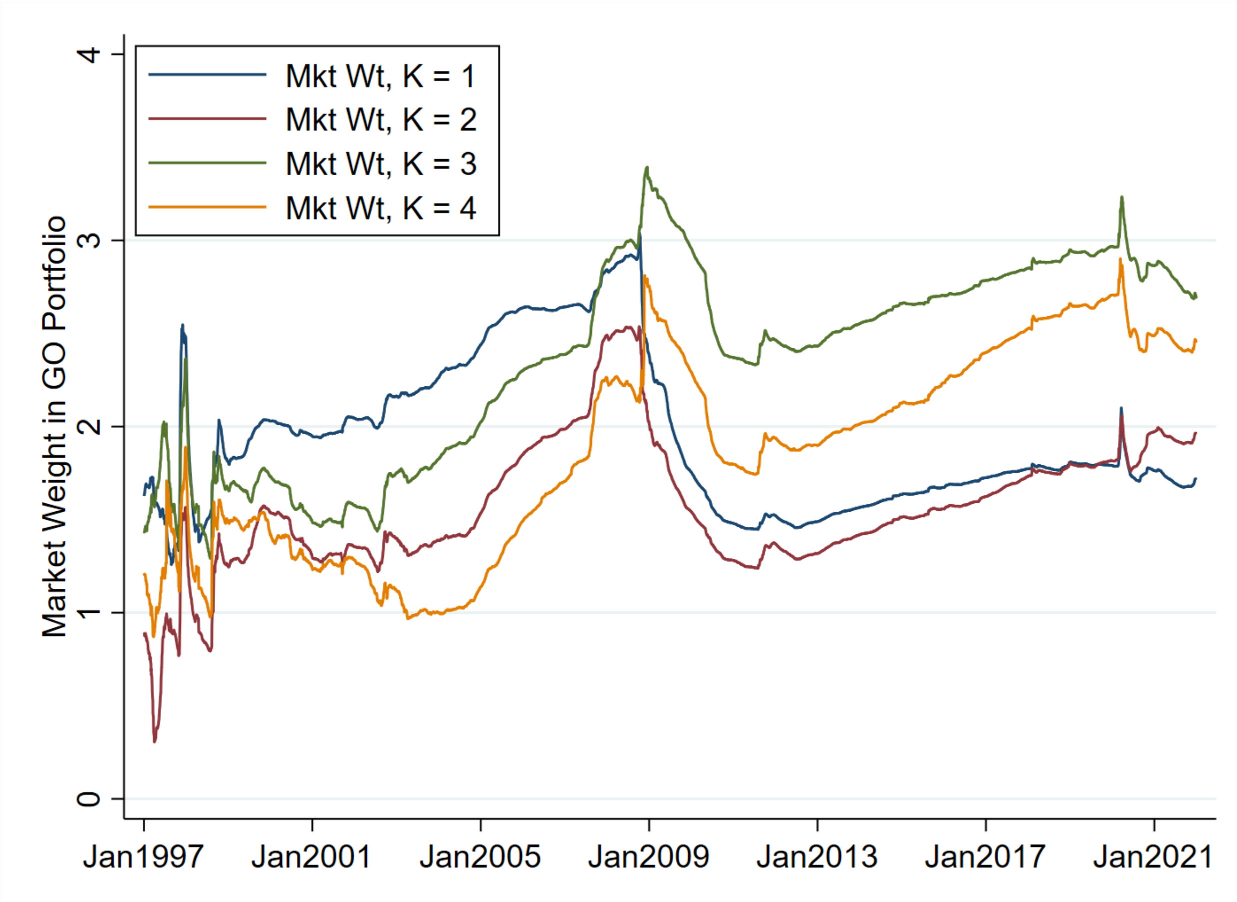
Panel B shows predicted variance from the fractionally integrated model and risk-neutral variance from option prices at the annual (360-day) horizon. The data are daily from 1996 through 2021.

Figure 1 (continued): The Estimated Variance Premium



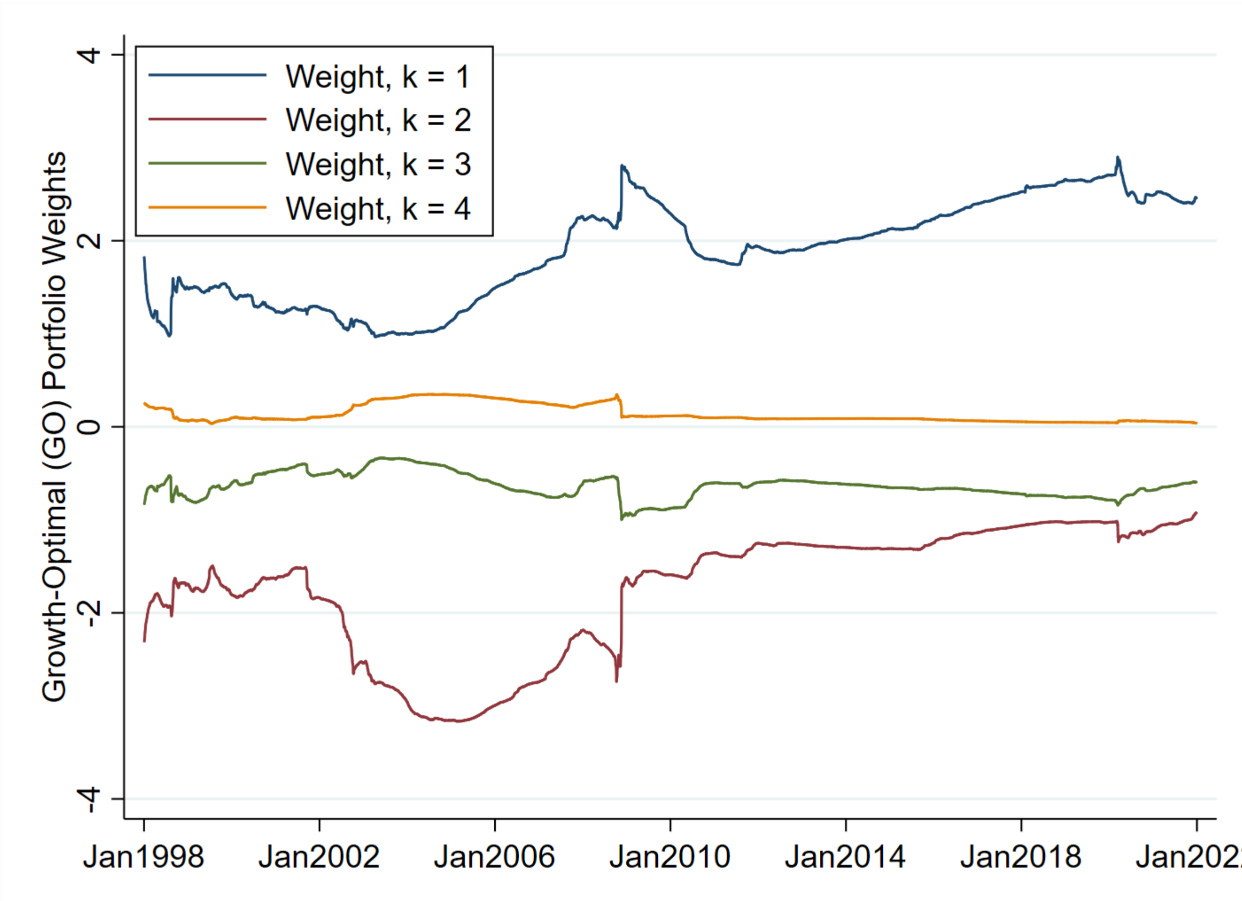
Panel C shows annualized variance premiums at the monthly (30-day) and annual (360-day) horizons. Each variance premium is the difference between risk-neutral variance and predicted variance from the fractionally integrated model at each horizon. The data are daily from 1996 through 2021.

Figure 2: Growth-Optimal Portfolio Weights



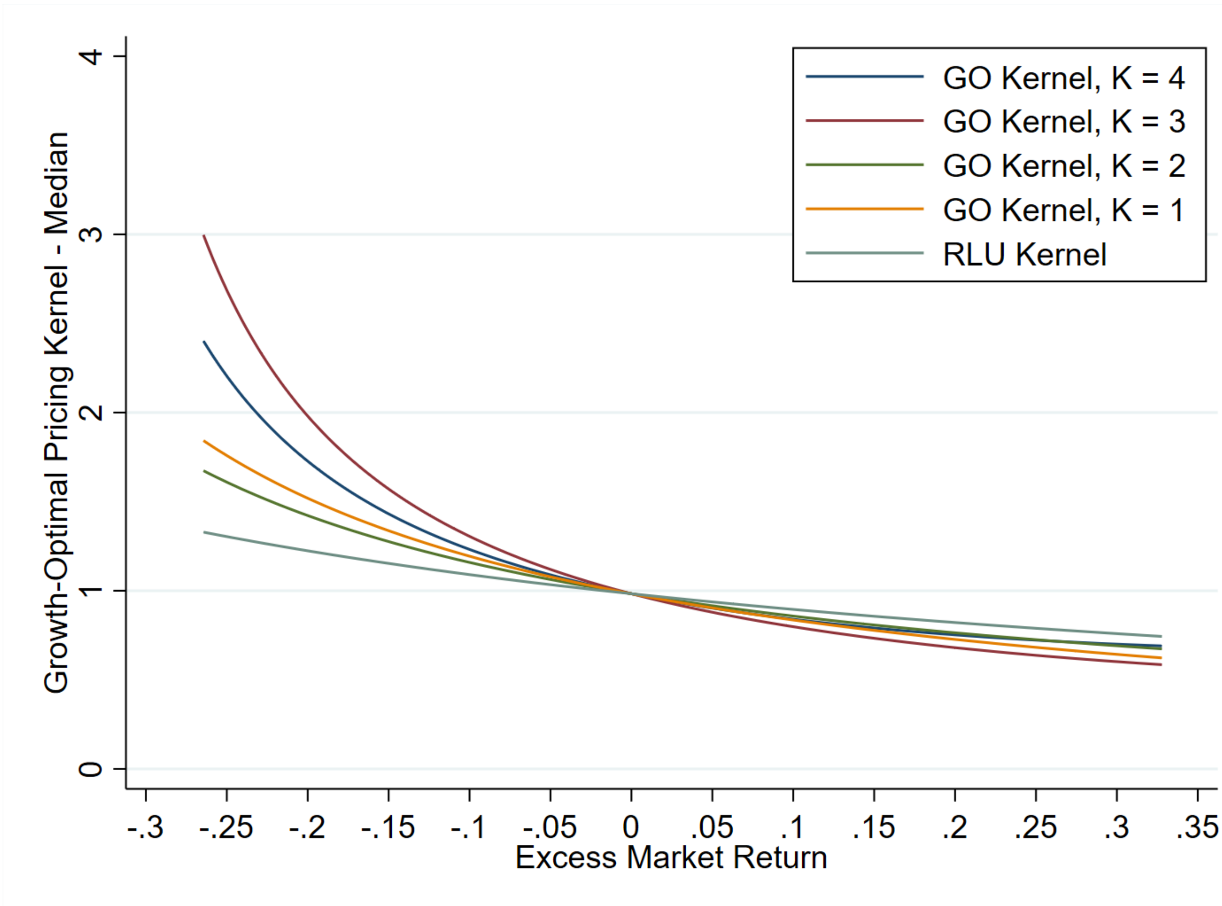
Panel A shows the weights on the stock market in the growth-optimal (GO) portfolio based on the risk-free asset, the stock market, and $K-1$ market-related securities. The market-related securities have excess returns equal to the market excess return raised to the k^{th} power, where $k=2,\ldots,K$. The figure shows the GO weight on the market for implied risk premium (IRP) models with $K=1,2,3,4$ risky assets. Weight estimates come from recursive regressions of the variance premium on K risk-neutral moments, as in equation (11). The data are daily from 1997 through 2021.

Figure 2 (continued): Growth-Optimal Portfolio Weights



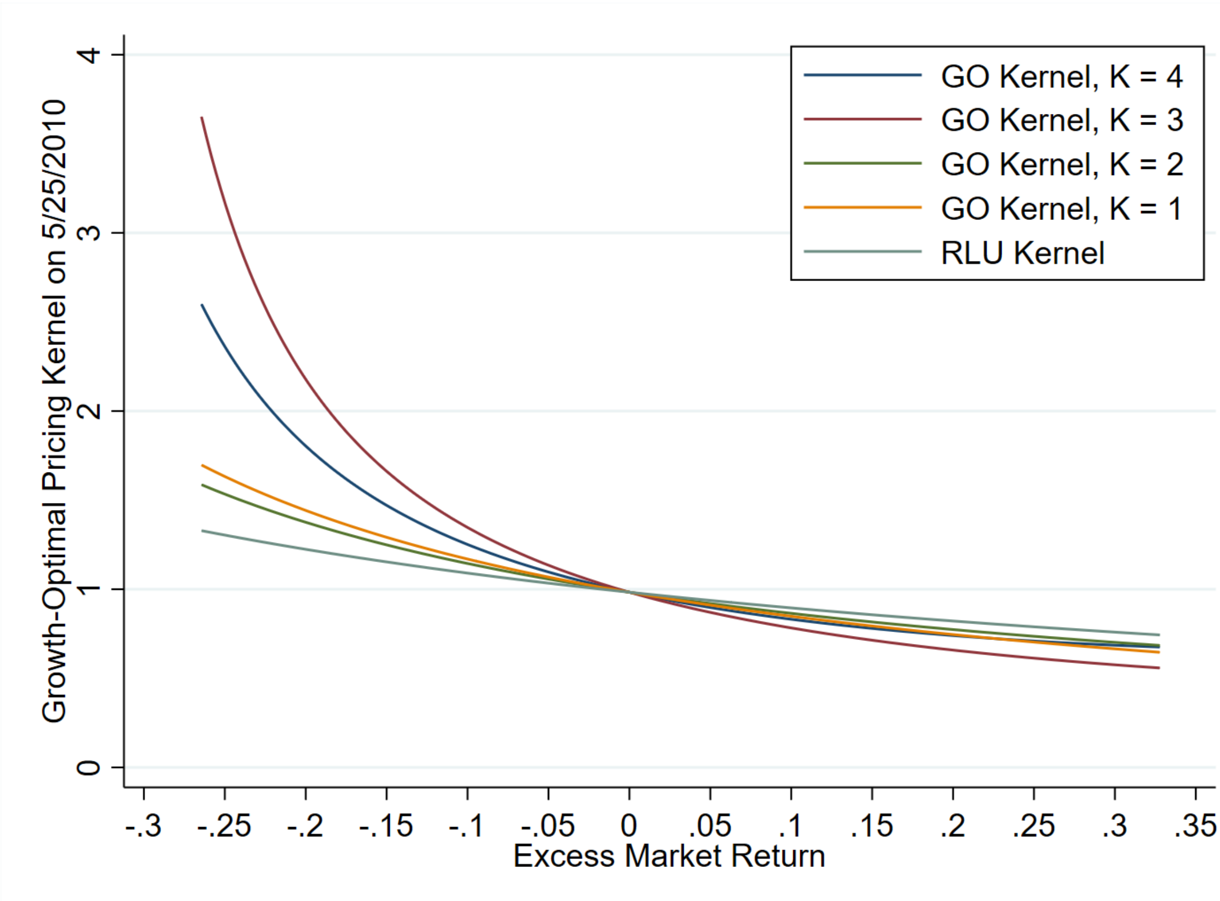
Panel B shows the weights on all ($K = 4$) risky assets in the growth-optimal portfolio based on the risk-free asset, the stock market ($k = 1$), and three market-related securities ($k = 2, 3, 4$). The market-related securities have excess returns equal the market excess returns squared ($k = 2$), cubed ($k = 3$), and raised to the fourth power ($k = 4$). Weight estimates come from recursive regressions of the variance premium on four risk-neutral moments with degrees three, four, five, and six, as in equation (11). The data are daily from 1998 through 2021.

Figure 3: Growth-Optimal Pricing Kernel



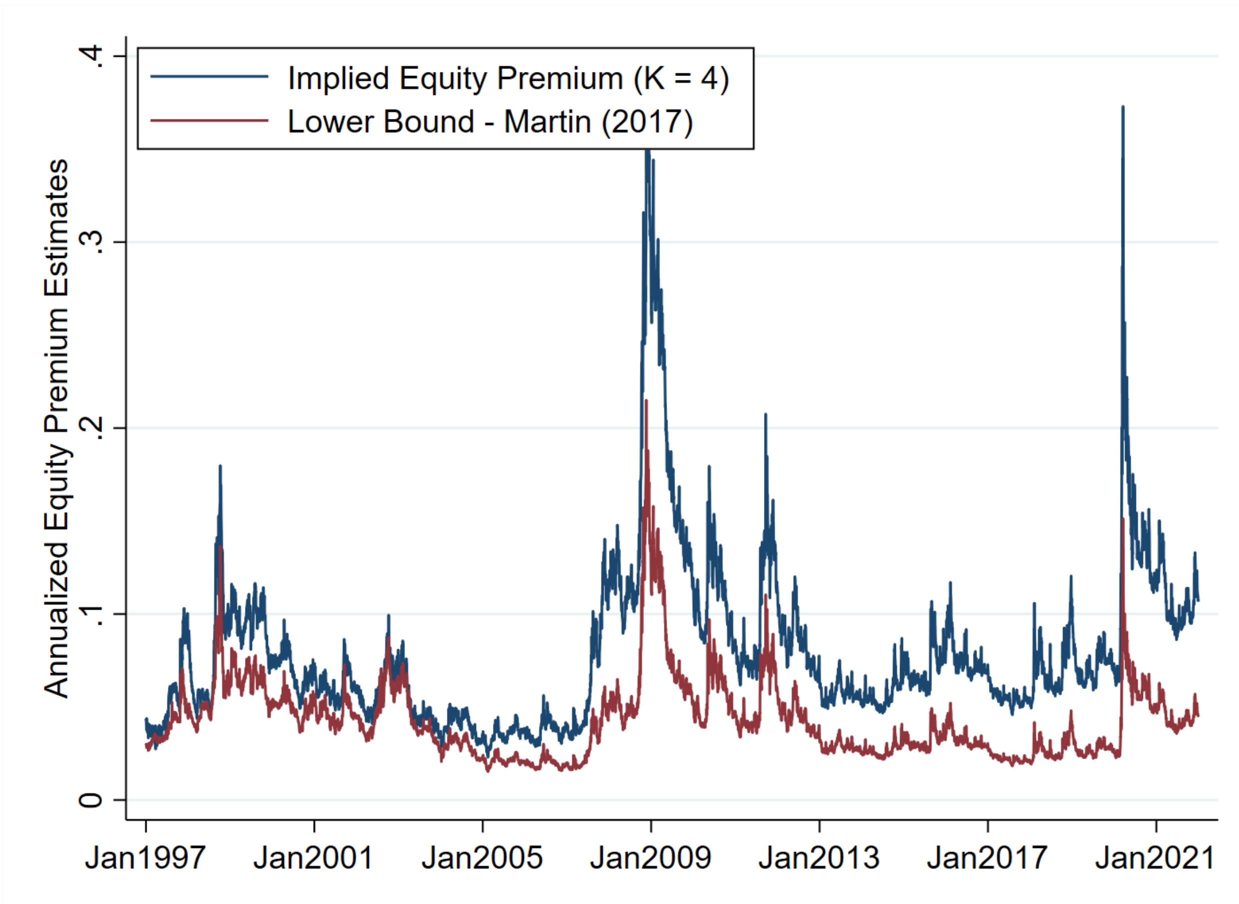
Panel A shows growth-optimal pricing kernels based on sample median weights on the risk-free asset, stock market, and $K-1$ market-related securities. The market-related securities have excess returns equal to the market excess return raised to the k^{th} power, where $k = 2, ..., K$. The figure shows the median GO pricing kernel for implied risk premium (IRP) models with $K = 1, 2, 3, 4$ risky assets. Weight estimates come from recursive regressions of the variance premium on K risk-neutral moments, as in equation (11). Median weights are based on daily data from 1997 through 2021. The representative log utility (RLU) kernel is a special case of the $K = 1$ kernel with a unity weight on the stock market and zero weights on other securities.

Figure 3 (continued): Growth-Optimal Pricing Kernel



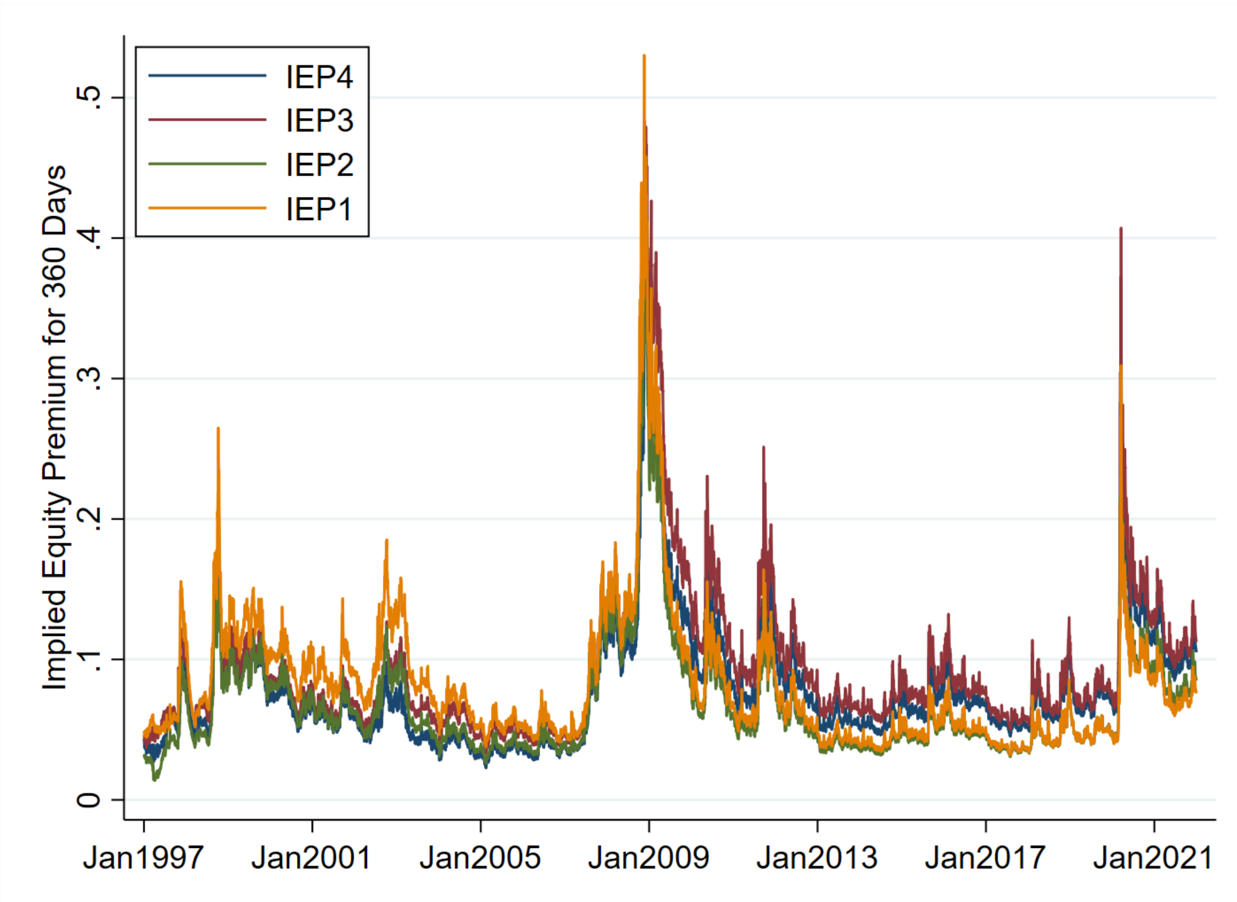
Panel B shows growth-optimal (GO) pricing kernels with weights on the risk-free asset, the stock market, and $K-1$ market-related securities on May 25, 2010. On this date, the monthly expected equity premium was at the 95th percentile for the sample. The market-related securities have excess returns equal to the market excess return raised to the k^{th} power, where $k = 2, ..., K$. The figure shows the GO pricing kernel for implied risk premium (IRP) models with $K = 1, 2, 3, 4$ risky assets. Weight estimates come from recursive regressions of the variance premium on four risk-neutral moments with degrees three, four, five, and six, as in equation (11). The representative log utility (RLU) kernel is a special case of the $K = 1$ kernel with a unity weight on the stock market and zero weights on other securities.

Figure 4: The Implied Equity Premium and Martin's Lower Bound



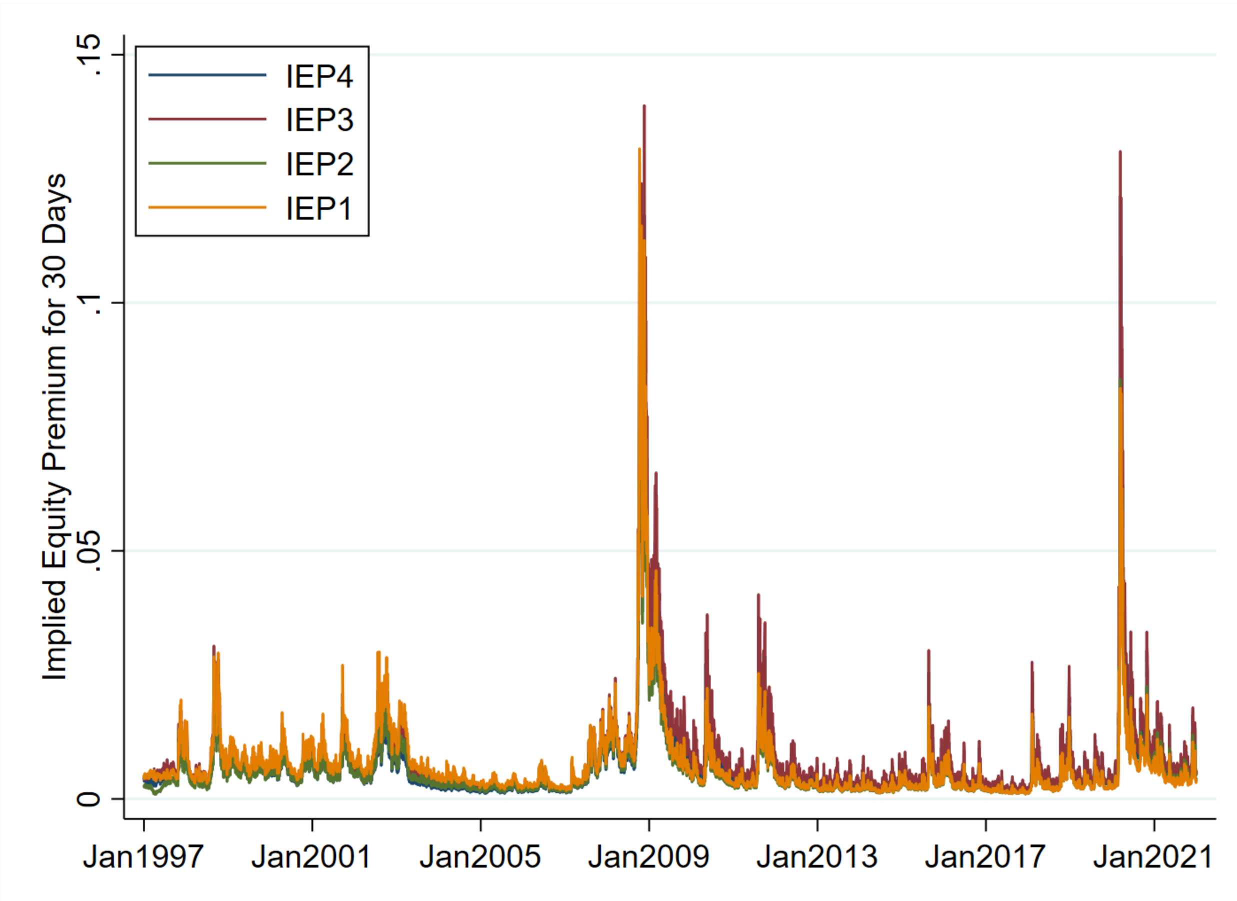
This figure shows the implied equity premium (IEP) and Martin's (2017) lower bound on the equity premium at the one-year (360-day) horizon from 1997 to 2021. The one-year IEP is based on the fourth-order ($K = 4$) approximation of the growth-optimal pricing kernel, IEP4_{360}, in equation (9). The lower bound is the representative log utility equity premium, RLUEP_{360}, which is a special case of equation (9) with $w_{1,t} = 1$ and $w_{k,t} = 0$ for $k > 1$. The data are daily from 1997 through 2021.

Figure 5: Implied Equity Premiums



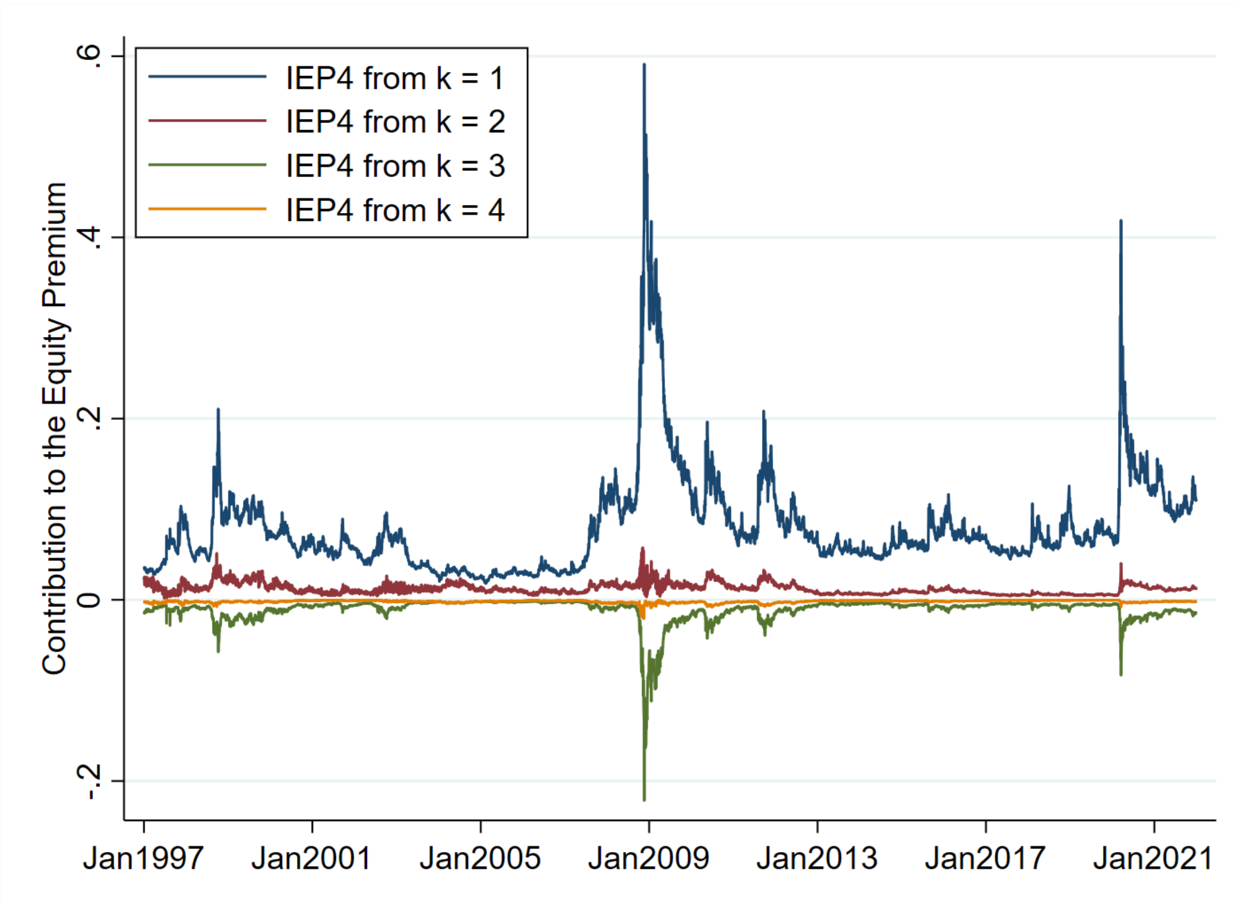
Panel A shows one-year (360-day) implied equity premiums (IEPs) for models that approximate the growth-optimal pricing kernel with degrees $K = 1, 2, 3, 4$. The data are daily from 1997 through 2021.

Figure 5 (continued): Implied Equity Premiums



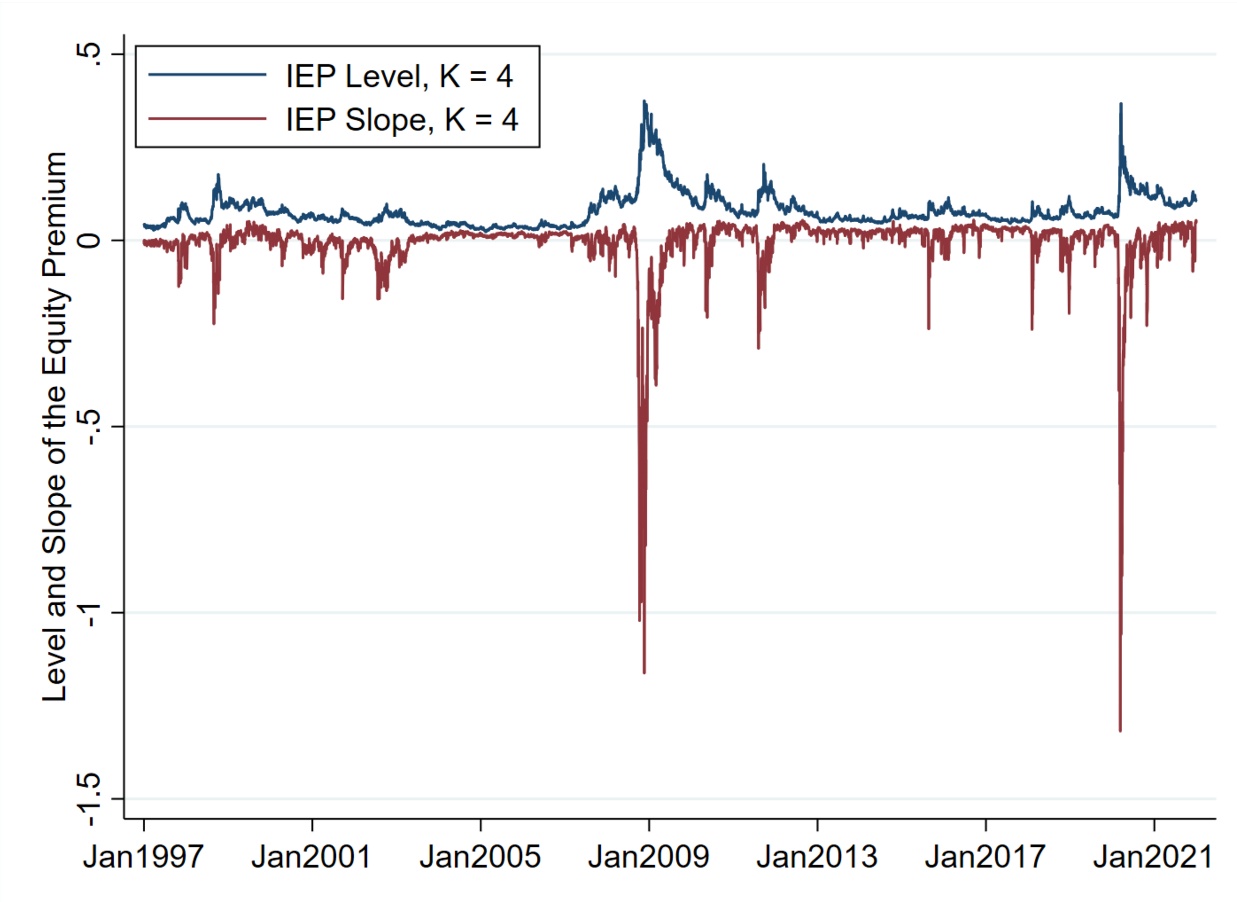
Panel B shows monthly (30-day) implied equity premiums (IEPs) for models that approximate the growth-optimal pricing kernel with degrees $K = 1, 2, 3, 4$. The data are daily from 1997 through 2021.

Figure 6: Decomposing the Implied Equity Premium



This figure shows components of the implied equity premium (IEP) at the one-year (360-day) horizon. The one-year IEP is based on the fourth-order ($K = 4$) approximation of the growth-optimal pricing kernel, IEP4_{360}, in equation (9). The components are the four terms ($k = 1, 2, 3, 4$) in equation (9), where k is the exponent applied to the market excess return. The data are daily from 1997 through 2021.

Figure 7: Term Structure of the Implied Equity Premium



This figure shows the level and slope of the term structure of the implied equity premium (IEP). The level is IEP4_{360} in equation (9) based on the fourth-order ($K = 4$) approximation of the growth-optimal pricing kernel. The slope is the difference between the annualized one-year and one-month IEPs divided by the difference in these maturities: $(365/360 \times \text{IEP4}_{360} - 365/30 \times \text{IEP4}_{30}) / ((360-30)/365)$. The data are daily from 1997 through 2021.

Appendix for
"The Implied Equity Premium"

by

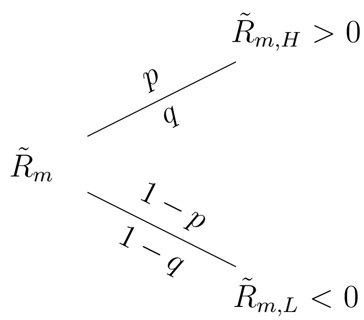
Paul C. Tetlock

Columbia Business School

A Recovery in a Two-State Model

In a simple setting, I show how to use empirical estimates of the variance premium and risk-neutral skewness to find the true equity premium and pricing kernel. As in the main model, available assets include a risk-free security with a gross rate of R_f, a stock market with an excess return of $\tilde{R}_m$, and options on the market with a continuum of strike prices. I assume that there are no arbitrage opportunities.

The model has two states: state H has high market excess returns, $\tilde{R}_{m,H} > 0$, and state L has low market excess returns, $\tilde{R}_{m,L} < 0$. State H occurs with probability $0 < p < 1$. These restrictions on return realizations and probabilities are necessary to ensure the absence of arbitrage. I parameterize return realizations so that the true equity premium is μ and return volatility is σ. This implies that $\tilde{R}_{m,H} = \mu + p^{-0.5}(1-p)^{0.5}\sigma$ and $\tilde{R}_{m,L} = \mu - p^{0.5}(1-p)^{-0.5}\sigma$. Negative excess returns in the low state implies an upper bound on the market Sharpe ratio: $\mu/\sigma < p^{0.5}(1-p)^{-0.5}$.



The econometrician does not observe the equity premium, μ, but has useful information about it. She observes the second moment of market excess returns, $\mathbb{E}\tilde{R}_m^2$, which is like empirical market variance. She effectively observes all risk-neutral moments of excess market returns, $\mathbb{E}^*\tilde{R}_m^j$ for $j = 1, ..., \infty$, because they are weighted sums of prices of market options. In this simple model, observation of risk-neutral variance and skewness, $\mathbb{E}^*\tilde{R}_m^2$ and $\mathbb{E}^*\tilde{R}_m^3$, will be sufficient to recover the physical return distribution.

By the no-arbitrage assumption, the growth-optimal portfolio consisting of the risk-free asset and stock market is the reciprocal of the pricing kernel, $m = \left(R_f + w\tilde{R}_m\right)^{-1}$, that prices these two assets. The two parameters are R_f, which is known, and w, which is the unknown weight on stocks. In this two-state model with no arbitrage, any pricing kernel that prices the risk-free rate and the stock market must also price all options based on the market because the former assets span the latter.

One can explicitly compute risk-neutral probabilities and pricing kernel values for the two states in terms of the unknown equity premium and market variance. Because the risk-neutral excess return of the market is zero, the risk-neutral probabilities of the high and low states must be $q = -\frac{\tilde{R}_{m,L}}{\tilde{R}_{m,H}+\tilde{R}_{m,L}} > 0$ and $1 - q = \frac{\tilde{R}_{H,L}}{\tilde{R}_{m,H}+\tilde{R}_{m,L}} > 0$, respectively. Because the pricing kernel is proportional to the ratio of risk-neutral to physical probabilities, its values in the high and low states must be $R_f \left(R_f + w\tilde{R}_{m,H}\right)^{-1} = q/p$ and $R_f \left(R_f + w\tilde{R}_{m,L}\right)^{-1} = (1-q)/(1-p)$, respectively. From the pricing kernel definition, this implies that the unknown weight on stocks is:

$$w = -\frac{R_f \mu}{\tilde{R}_{m,L}\tilde{R}_{m,H}}.$$

This weight on stocks is positive, $w > 0$, because $-\tilde{R}_{m,L} > 0$.

I now show how to estimate w and thus μ from option prices and physical return variance. The first step is to express observable quantities in terms of the unknown parameters of the physical return distribution. Risk-neutral variance is:

$$\begin{aligned}
\mathbb{E}^* \tilde{R}_m^2 &= q\tilde{R}_{m,H}^2 + (1-q)\tilde{R}_{m,L}^2 \\
&= -\tilde{R}_{m,L}\tilde{R}_{m,H},
\end{aligned}$$

which is positive because $-\tilde{R}_{m,L} > 0$. Risk-neutral skewness is:

$$\begin{aligned}
\mathbb{E}^* \tilde{R}_m^3 &= q\tilde{R}_{m,H}^3 + (1-q)\tilde{R}_{m,L}^3 \\
&= -\tilde{R}_{m,L}\tilde{R}_{m,H}\left(\tilde{R}_{m,H} + \tilde{R}_{m,L}\right).
\end{aligned}$$

The variance premium is:

$$\begin{aligned}
\mathbb{E}^* \tilde{R}_m^2 - \mathbb{E}\tilde{R}_m^2 &= -\tilde{R}_{m,L}\tilde{R}_{m,H} - \left(\sigma^2 + \mu^2\right) \\
&= \left[p^{0.5}(1-p)^{-0.5} - p^{-0.5}(1-p)^{0.5}\right]\mu\sigma - 2\mu^2.
\end{aligned}$$

The second step is applying the variance premium equation (11) in the main model to the case here with a single parameter, w, in the pricing kernel and solving for w to obtain:

$$\begin{aligned} w &= \frac{\mathbb{E}_t^* \tilde{R}_m^2 - \mathbb{E}\tilde{R}_m^2}{-R_f^{-1}\mathbb{E}_t^* \tilde{R}_m^3} \\ &= -\frac{R_f \mu}{\tilde{R}_{m,L}\tilde{R}_{m,H}}, \end{aligned}$$

which is the same as the true value for w.

The third and last step is applying the equity premium equation (10) in the main model to the linear pricing kernel with slope w to obtain:

$$\begin{aligned} \mathbb{E}\tilde{R}_m &= R_f^{-1} w \mathbb{E}^* \tilde{R}_m^2 \\ &= R_f^{-1} \frac{R_f \mu}{\tilde{R}_{m,L}\tilde{R}_{m,H}} \left(\tilde{R}_{m,L}\tilde{R}_{m,H}\right) \\ &= \mu. \end{aligned}$$

This last equation shows that the option-implied equity premium equals the true equity premium in this simple model.

Recovery of the true pricing kernel, m, and equity premium, μ, is exact for any physical distribution parameters (p, μ, σ) that satisfy the no-arbitrage assumption. In particular, the recovery procedure works regardless of the signs of the variance premium and risk-neutral skewness. The variance premium is positive whenever $p > 0.5 + 0.5\left[\mu^2\left(\mu^2+\sigma^2\right)^{-1}\right]^{0.5}$ or $p < 0.5 - 0.5\left[\mu^2\left(\mu^2+\sigma^2\right)^{-1}\right]^{0.5}$. The former condition is satisfied for reasonable equity premium and volatility values, such as $\mu = 0.05$ and $\sigma = 0.15$, because stocks outperform Treasuries in most years ($p \approx 2/3$). The conditions for negative risk-neutral skewness are the same as those for a positive variance premium. Since these two quantities always have opposite signs, the pricing kernel parameter, w, which is the negative of the ratio of these quantities, is always positive. In the empirically relevant case, the variance premium is positive, risk-neutral skewness is negative, and w is positive.

B Measuring Option Prices and Market Variance

B.1 Option Prices

I compute risk-neutral moments of market returns from SPX option prices using OptionMetrics data. I adopt many data filtering methods from Chang et al. (2013) and Martin (2017) to eliminate securities with low liquidity and unreliable prices. I augment these methods using a customized no-arbitrage filter for options with extreme strike prices, which are important for higher-order risk-neutral moments.

As in Martin (2017), I remove SPX options with:

- maturities less than 7 days or greater than 549 days
- duplicated data
- p.m. settlement
- non-positive closing bid
- quarter-end expiration dates
- non-null expiration indicator
- arbitrage violations relative to the SPX index
- higher midpoint price among put and call for each date, maturity, and strike price.

The minimum price filter restricts the sample to out-of-the-money options. For each maturity, I set the SPX futures price as the lowest strike price such that an SPX put exceeds the value of an SPX call by one penny.[20] I also drop options with open interest less than 1000 contracts to ensure sufficient liquidity. I drop options with zero or missing implied volatility, which violate arbitrage bounds.

[20]This definition is almost perfectly correlated with the futures price implied by futures-spot parity.

I drop a few options with glaring data errors to reduce outliers, though these omissions have little substantive impact on the results. Based on sudden changes in data coverage and extreme illiquidity, I drop all options with the following seven date-maturity combinations and nine contracts with the following date-maturity-strike-type combinations:

Date	Maturity	Strike	Type
1998-03-24	1999-03-20	All	Both
1998-09-21	1999-09-18	All	Both
1999-09-20	2000-09-16	All	Both
1999-09-21	2000-09-16	All	Both
1999-09-22	2000-09-16	All	Both
2007-06-19	2008-06-21	All	Both
2001-09-17	2002-06-22	1900	Call
2001-09-18	2002-06-22	1900	Call
2001-09-19	2002-06-22	1900	Call
2001-09-20	2002-06-22	1600	Call
2001-09-20	2002-06-22	750	Put
2001-09-21	2002-06-22	800	Put
2006-01-19	2006-12-16	1900	Call
2008-10-10	2009-09-19	1525	Call
2011-06-06	2012-06-16	3000	Call

I design a set of filters to ensure sufficient information for the computation of risk-neutral moments at each horizon. For each date-maturity pair, I require at least:

- a moneyness (i.e., strike/index) range from 95% to 105%
- two calls
- two puts
- five options of any type

The minimum moneyness range drops 0.8% of date-maturity pairs and the last three requirements drop only 0.1% of pairs.

Whereas Chang et al. (2013) drop options with bid-ask midpoints of less than \$0.375 (3/8), I drop options with extreme moneyness until options with adjacent prices differ by at least a penny. This filter eliminates most remaining arbitrage violations and illiquid options.

B.2 Risk-Neutral Moments

As in Chang et al. (2013) and Martin (2017), I compute risk-neutral moments based on observed option prices and strike prices for each date and maturity. I use a discrete approximation of the integral in equation (20) based on the range of available strike prices supplemented using extrapolated option prices, following Chang et al. (2013). I assume options with strikes outside the available range of moneyness have implied volatility equal to a trend-line extrapolation of the volatility skew based on the three options with the closest strike prices. I estimate the trend-line parameters by minimizing the sum of absolute deviations from the three closest strikes. I constrain the absolute value of the slope to be no greater than 1.0, which affects almost no data. I only extrapolate insofar as the range of available moneyness is less than two standard deviations above and below the futures price. I also winsorize the implied volatility range at 5% and 100% and the moneyness range at 1% and 300% of the futures price, though these limits rarely bind. This extrapolation procedure creates an additional 1.8% of options beyond the original cleaned data.

As in Martin (2017) and related studies, I must interpolate between the irregular maturities of the resulting risk-neutral moments to obtain the desired uniform maturities of monthly, bimonthly, quarterly, and annual because option expiration dates occur at fixed times within months. I design an interpolation procedure to maximize data availability and minimize likely errors. First, I set the raw odd (even) risk-neutral moments to missing for date-maturity pairs with non-negative (non-positive) values, which would be inconsistent with theory and prior evidence. This filter only affects 45 third-order moments (0.1%) and 182 fifth-order moments (0.5%) out of 51,013 date-maturity pairs; and it affects no even moments. I also set a handful of third- and fifth-order moments that exceed a very small negative (annualized) value, -0.0001, to missing.

Second, for each moment, date, and irregular maturity, I fill in missing daily moments using adjusted lagged weekly moments. To this end, I partition all moments into six maturity groups with cutoffs defined by the uniform maturities of 30, 60, 90, 180, and 360 days, creating groupings of [7,30], [31, 60], [61, 90], [91, 180], [181, 360], and [361, 549] days. For each group, date, and moment, I estimate the lagged weekly moment using a rolling median of the moment's non-missing values on days -7 to -1. I fill in missing weekly medians for the longest horizons, e.g., [361, 549], using values from the next longest horizon, e.g., [181, 360]. I then apply this procedure to the shortest horizons with missing weekly data using the next shortest horizon. These missing data procedures apply to fewer than 0.7% of observations for all weekly moments.

Third, to obtain values for standardized maturities, I interpolate between adjacent horizon groups based on the local slope of the term structure of the risk-neutral moment. For example, I estimate the third risk-neutral moment at a 30-day maturity using a [1/3, 2/3] weighted average of the third risk-neutral moments with 10-day and 40-day maturities. I apply this interpolation to the raw daily moments and to the lagged weekly moments, filling in any remaining missing values of weekly moments with most recently available values. Lastly, I fill in any missing daily moments using adjusted weekly moments. The adjusted weekly moment is the unadjusted weekly moment plus the median difference between the daily and weekly moments across all horizons for which both are nonmissing. I apply this same interpolation procedure to OptionMetrics' risk-free rates with irregular maturities to obtain risk-free rates with standardized maturities.

B.3 ETF Variance

The RV estimator is the sum of squared log SPY ETF returns over 78 intraday intervals, where 78 is the number of 5-minute periods in regular exchange trading hours. The final RV estimator is the average of 10 sub-sampled RV estimators based on 10 staggered sets of 78 non-overlapping intervals. Each set of 78 return intervals is based on 79 trade prices that are equally spaced in business time—i.e., equal numbers of trades—throughout regular trading hours. Thus, the 10 RV estimators use a total of 790 prices equally spaced in business time. The first RV estimator uses prices 1, 11, 21, ..., 781, the second uses prices 2, 12, 22, ..., 782, and the tenth uses prices 10, 20, 30, ..., 790.

For these estimators, I use only trades that satisfy the following standard conditions, applied in order below:

- price is positive
- quantity is positive
- time must be between 9:30 a.m. and 4:00 p.m. Eastern time
- correction code is "00"
- condition code is not 1, 4, 7, 8, 9, A through D, G, H, K, L, N, P, R, S, U through W, Y, or Z

- exchange is the most active for SPY on that day
- price must be within the range of the ETF's daily low and high

I aggregate all trades by transaction time and use the median trade price for each time as the price. I keep only trades from the most active exchange on each day. I drop trades with extreme reversals as defined an absolute return exceeding five times the 50-observation rolling average. The number of unique transaction times is the basis for the partition into 790 prices equally spaced in business time.

I fix two glaring data errors in trade prices. I obtain the daily low and high prices from the Center for Research in Securities Prices (CRSP) database. On March 31, 1997, I use high and low prices from Yahoo! Finance because the CRSP high and low prices are incorrect. On May 12, 1999, I set SPY prices in TAQ that are below the daily CRSP low to the index low for trades before 12 noon and to the previous valid trade price for trades after noon. These changes improve aesthetics by reducing outliers but have no substantive impact on the results.