Higher-order Risk Premium, Stock Return Predictability, and Rare Event Dynamics *

Zhenzhen Fan†  Xiao Xiao‡  Hao Zhou§

January 18, 2019

Abstract

We find that higher-order risk premium as a component of the total variance risk premium can significantly predict market returns over medium-term horizon. Decomposing the total variance risk premium improves return predictability both in-sample and out-of-sample, with $R^2$ up to 14 percent for the 3-month horizon. This finding proves to be economically significant in an asset allocation exercise, becomes even stronger for the portfolio returns of the momentum factor, and survives a series of robustness checks. We show that a consumption-based asset pricing model with rare events can generate the predictability from the higher order risk premium.

JEL Classification: G12, G13, C22

Keywords: Equity risk premium; Predictive regression; Variance risk premium; Higher order risk premium; Option implied information; Rare events

---

*We thank Bakshi Gurdip, Bing Han, Anthony Neuberger, Cisil Sarisoy, Paul Schneider, George Tauchen, Fabio Trojani, Grigory Vilkov, Dacheng Xiu, and Lai Xu for helpful comments. We would also like to thank seminar participants at Erasmus University Rotterdam, European Finance Association annual conference and Asian Finance Association annual conference. All remaining errors are our own.

†School of Finance, Nankai University. Email: Zhenzhen.fan@nankai.edu.cn.

‡Erasmus School of Economics, Erasmus University Rotterdam. Email: xiao@ese.eur.nl.

§PBC School of Finance, Tsinghua University. Email: zhouh@pbc Beijing. edu.cn.
Higher-order Risk Premium, Stock Return Predictability, and Rare Event Dynamics

This Version: November 2018

Abstract

We find that higher-order risk premium as a component of the total variance risk premium can significantly predict market returns over medium-term horizon. Decomposing the total variance risk premium improves return predictability both in-sample and out-of-sample, with $R^2$ up to 14 percent for the 3-month horizon. This finding proves to be economically significant in an asset allocation exercise, becomes even stronger for the portfolio returns of the momentum factor, and survives a series of robustness checks. We show that a consumption-based asset pricing model with rare events can generate the predictability from the higher order risk premium.

**JEL Classification:** G12, G13, C22

**Keywords:** Equity risk premium; Predictive regression; Variance risk premium; Higher order risk premium; Option implied information; Rare events
1 Introduction

The conventional variance risk premium—defined as the difference between the risk-neutral and realized return variances—proves to be a strong predictor for the aggregate market returns in the short run (see, e.g., Bollerslev, Tauchen, and Zhou (2009), Drechsler and Yaron (2010), and Bekaert and Hoerova (2014)). Alternatively, recent studies show that various jump and tail risk measures have return predictability at medium-term horizons, beyond the information contained in the conventional variance risk premium (see, e.g., Du and Kapadia (2012), Bollerslev, Todorov, and Xu (2015), and Guo, Sha, Wang, and Zhou (2018)). In this paper, we incorporate these two types of predictability evidence in an integrated and coherent decomposition framework.

We decompose the total variance risk premium (TVRP) into a pure second order, variance risk premium (VRP) and a higher-order risk premium (HRP), in a model-free fashion. Using the insights from Neuberger (2012), Kozhan, Neuberger, and Schneider (2013), and Martin (2017), one can replicate the higher moments of returns under the risk-neutral measure with the portfolios of out-of-the-money European call and put options. Matching the risk-neutral moments with their realized counterparts, we can decompose the TVRP into the premiums associated with the second and higher moments. Various ad hoc jump and tail risk measures considered in the literature are closely related to the option-implied or realized higher moments in our setting.

The two components of TVRP have their own unique empirical features and distinct economic interpretations. Using S&P 500 index and its option data from 1996 to 2016, we find that the VRP is on average positive as documented in literature. The average HRP, mainly attributed to the risk premium related to the third moment of returns, has the opposite sign. HRP might reflect compensation for unexpected large and discontinuous movement of stock returns, while VRP could represent compensation for uncertain continuous and diffusive movement of stock return variance. As a consequence, aggregating VRP and HRP may lead to substantial information losses, and the return predictability afforded by TVRP may be enhanced by considering these two components separately.
We find that the decomposition improves both in-sample and out-of-sample predictability of TVRP over 1- to 24-month horizons. In particular, the adjusted $R^2$'s of the joint regressions with VRP and HRP are 14.0%, 13.0% and 6.6% at 3-, 6-, and 12-month horizons, in contrast to 8.0%, 2.7% and -0.3% of the univariate regressions with TVRP alone. The out-of-sample $R^2$'s of the joint regressions with VRP and HRP are 14.5%, 11.0% and 3.6% at 3-, 6-, and 12-month horizons, compared with 7.7%, -10.1% and −11.46% in the univariate regressions with TVRP alone. In univariate regressions, VRP contains short-term predictability with significant positive coefficients, while HRP contains medium-term predictability with significant negative coefficients. At 1- and 3-month (9- to 24-month) horizons, VRP (HRP) has higher $R^2$'s than TVRP, both in-sample and out-of-sample. The ad hoc decomposition approaches in Du and Kapadia (2012), Bollerslev, Todorov, and Xu (2015), and Guo, Sha, Wang, and Zhou (2018) could not produce in-sample $R^2$'s as high as ours or positive out-of-sample $R^2$’s for all horizons.

We further examine the economic significance of the predictability offered by the two components of TVRP. First, our finding on HRP seems to survive the exclusion the global financial crisis period—a notable rare disaster realized in our sample, indicative of a structural explanation. Second, we examine the economic significance of the two components of the TVRP via an asset allocation experiment, which generates higher out-of-sample utility gains than that generated by TVRP alone. Finally, we examine the predictability of the two components of TVRP for the returns of size, value, and momentum portfolios, with even higher increases in return predictability than TVRP alone for the momentum portfolio. We also conduct extensive robustness analysis. The significant predictability afforded by VRP and HRP remains robust after controlling for the ten variables in Welch and Goyal (2008) and the short interest in Rapach, Ringgenberg, and Zhou (2016). Our main results hold when we calculate TVRP, VRP, and HRP using expected variances estimated from forecasting regressions and realized variance estimated from high-frequency 5-minute data. The prediction results are robust to alternative moments calculations as in Bakshi, Kapadia, and Madan (2003) and Schneider and Trojani (2017).

To understand the economic mechanism behind the complementary predictability
from VRP and HRP, we propose a Lucas-type equilibrium model with stochastic volatility-of-volatility and time-varying rare events in the consumption dynamics. Our model endogenously generates positive and negative predictive coefficients on VRP and HRP, respectively, implying that aggregating these components into TVRP conflates the two opposite effects. Moreover, we show that, while the short term predictability from VRP can be explained by the volatility-of-volatility process with moderate persistence, the medium term predictability in HRP comes from both the more persistent intensity of rare event and the capability of HRP to predict rare events in future consumption dynamics. Our model is able to reproduce the differential patterns of the predictive $R^2$'s from VRP, HRP, and TVRP in the observed sample.

**Literature**

Our paper contributes to the literature on decomposing the conventional variance risk premium. Todorov (2009) analyses the pricing of two components of variance risk due to stochastic volatility and jumps in a semiparametric framework. Bollerslev, Todorov, and Xu (2015) provide an indirect decomposition of the conventional variance risk premium by extracting the component associated with a tail index. Their decomposition involves approximation and estimation of the tail shape parameter and jump intensity parameter in an optimization procedure. Our approach differs from theirs in two important aspects. First, the conventional variance risk premium essentially contains information of not only the second, but also the third and higher moments of return premium. Our decomposition is therefore internally consistent and coherent. Second, our decomposition is completely model-free and does not require any econometric estimation, ensuring that the decomposed premiums are linearly replicable. The replication of all higher moments are guaranteed under very general conditions in the discrete sample, including the circumstances when the market is incomplete or the price can jump.

Our paper also contributes to the literature on return predictability from the variance risk premium and jump or tail risk measures. Bollerslev, Tauchen, and Zhou (2009), Drechsler and Yaron (2010), and Bekaert and Hoerova (2014) show that the variance risk premium predicts aggregate market returns for up to a few months horizon. Du and Ka-
Padia (2012) find that model-free jump or tail index provide medium-term predictability for market returns. Bollerslev, Todorov, and Xu (2015) show that the extracted left jump tail has stronger return predictability than previously shown. Guo, Sha, Wang, and Zhou (2018) find that the realized signed jump risks predict short-horizon excess stock returns and economic fundamentals. Our paper reconciles these two types of predictability evidence by showing that the predictability offered by various jump and tail measures can be synthesized by the HRP embedded in the TVRP. Feunou, Jahan-Parvar, and Okou (2017) study the predictability of the downside variance risk premium, while Kilic and Shaliastovich (2017) show that the good and bad variance risk premiums can jointly predict stock and bond returns—both approaches are conceptually tied to the skewness premium in our decomposition framework.

Finally, our paper is related to the literature on rare events. Existing studies acknowledge the importance of rare events in explaining many economic puzzles. For example, Rietz (1988) introduces a low probability crash state to the two-state Markov-chain model to explain the equity-premium puzzle. This model is later extended by Barro (2006) and Wachter (2013). Liu, Pan, and Wang (2004) demonstrate in an equilibrium pure-exchange economy that uncertainty aversion towards rare events could explain index option smirks. Gabaix (2012) solves ten puzzles in macro-finance by incorporating a time-varying severity of disasters. This paper shows that rare events also give rise to the predictability of the market return from HRP—the higher order risk premium, as opposed to VRP—the variance risk premium, over medium-term forecasting horizons.

The rest of the paper is organized as follows. Section 2 develops the methodology to decompose TVRP into VRP and HRP. Section 3 explains the data used in the empirical analysis. Section 4 reports the joint predictive regression results for the aggregate market return from VRP and HRP, along with analysis on economic significance. Section 5 presents a series of statistical robustness checks. Section 6 proposes a general equilibrium model with rare events to explain the complementary predictability from VRP and HRP. Section 7 concludes.
2 Decomposing the Total Variance Risk Premium

In this section, we show how to identify the second, third, and fourth order risk-neutral moments, using linear combinations of $VIX^2$, $SVIX^2$ proposed by Martin (2017), and a skewness measure $KNS$ proposed by Kozhan, Neuberger, and Schneider (2013). Matching the risk-neutral components with their realized counterparts, we decompose the total variance risk premium into a second order risk premium or variance risk premium, and a higher order risk premium.

2.1 Total Variance Risk Premium

The total variance risk premium (TVRP) is the difference between $VIX^2$ and the total realized variance, where the total realized variance is defined as $2(e^r − 1 − r)$ with $r$ being the log return of the futures price. The advantage of this definition, compared with the conventional variance risk premium, which uses the realized variance of log returns, lies in the internal consistency between the option-implied moments and their realized counterparts. This facilitates a disaggregation of the total variance risk premium into components associated with different order moment risks.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space and $F_t(T)$ be the futures price of the asset at time $t$ with maturity $T > t$. Throughout this paper, we work with futures price instead of spot price to avoid complications with interest rates and dividends. We fix the time-to-maturity $\tau$, with $\tau = T − t$, and denote $F_t$ instead of $F_t(T)$ when no confusion is caused.

We define $g(r(t, T))$ as a twice-differentiable function, where the argument $r(t, T)$ is the holding period log return of the futures between time $t$ and $T$, i.e., $r(t, T) := \log(F_T) − \log(F_t)$. If there is no arbitrage opportunity in the market, the payoff $g(r(t, T))$ can be statically replicated by a continuum of options by applying the approach of Bakshi and Madan (2000) under very general conditions. The static replication strategy gives the expectation of $g(r(t, T))$ under the risk-neutral measure $Q$. In this paper, we construct payoffs that are power functions of returns. The expectation of such payoffs under the
risk-neutral measure $Q$ is the implied moment of returns.

We first consider the following payoff function,

$$g_v(r) := 2(e^r - 1 - r).$$

(1)

As shown in Bondarenko (2014), the expectation of $g_v(r)$ is a form of generalized variance, because $g_v(r)$ differs from $r^2$ only in higher order terms. To see this, we apply Taylor expansion to Equation (1) and get

$$g_v(r(t, T)) = r(t, T)^2 + \frac{1}{3} r(t, T)^3 + \frac{1}{12} r(t, T)^4 + o(r(t, T)^4).$$

(2)

While the payoff function $g_v$ may not look familiar, Kozhan, Neuberger, and Schneider (2013) show that its expectation under the risk-neutral measure $Q$ is the continuous version of the CBOE VIX index:

$$\text{VIX}_t^2 = \frac{1}{T-t} \mathbb{E}_t^Q \left[ g_v(r(t, T)) \right] = \frac{2}{T-t} \int_0^\infty \frac{\Theta_t(K, T)}{K^2} dK,$$

(3)

where $\Theta_t(K, T)$ denotes the time-$t$ value of an out-of-the-money option with strike price $K > 0$ and maturity $T$. Note that we do not need $F_t$ to be continuous for Equation (3) to hold, whereas using $\text{VIX}^2$ as a risk-neutral expectation of the quadratic variation of log returns requires the assumption of continuous price path.\footnote{Carr and Wu (2009) show that under general jump-diffusion price dynamics, the quadratic variation on the futures return from time $t$ to $T$ is given by}

$$2 \int_0^\infty \frac{\Theta_t(K, T)}{K^2} dK + \epsilon,$$

where $\epsilon$ is the approximation error and a function of jumps in the futures price. If the futures price can jump, the squared VIX index is not a consistent estimator of the risk-neutral expectation of the quadratic variation of log returns.

\footnote{Carr and Wu (2009) show that under general jump-diffusion price dynamics, the quadratic variation on the futures return from time $t$ to $T$ is given by}
as \( r_i \) for simplicity. The realized total variance, RTV, is then defined as,

\[
\text{RTV}_T := \frac{1}{T-t} \sum_{i=1}^{n} 2(e^{r_i} - 1 - r_i).
\]

(4)

To be consistent with the recent literature (e.g. Bollerslev, Tauchen, and Zhou (2009), Buss, Schoenleber, and Vilkov (2017), etc.), we define the total variance risk premium (TVRP) as the difference between ex ante risk-neutral expectation of the future return variation over \([t, t + \tau]\) and the ex post realized total variance over \([t - \tau, t]\),

\[
\text{TVRP}_t := \text{VIX}_t^2 - \text{RTV}_t.
\]

(5)

This formulation has the advantage that both implied and realized moments are available at time \( t \), facilitating the return prediction exercise in Section 4.\(^2\)

Many papers in the literature use different formulations of realized variance other than the realized total variance in Equation (4) to define TVRP. For instance, Carr and Wu (2009) use the realized squared simple returns, and Bollerslev, Tauchen, and Zhou (2009) use the realized squared log returns. However, since the squared VIX is the risk neutral expectation of the payoff \( g_v \), RTV serves as the only coherent realized counterpart of \( \text{VIX}^2 \).

### 2.2 Decomposition of TVRP

As shown in Equation (2), although \( g_v \) is primarily driven by the second order of returns, it is also exposed to higher orders of returns. In this section, we decompose \( g_v \), and subsequently, TVRP, into components related to the second and higher orders of returns. Similar to the expectation of \( g_v(r) \), the decomposed components can all be replicated using option portfolios in a model-free fashion.

Inspired by Martin (2017) and Kozhan, Neuberger, and Schneider (2013), we consider

\(^2\)In Section 5.2, we use forecasting models to predict one-period-ahead realized variance between \([t, t + \tau]\).
the simple variance payoff function \( g_{sv}(r) \) and a payoff function related to skewness \( g_{kns}(r) \), defined respectively as,

\[
g_{sv}(r) := (e^r - 1)^2, \quad g_{kns}(r) := 6(2 + r - 2e^r + re^r).
\]  

Under the general condition of no arbitrage, Martin (2017) and Kozhan, Neuberger, and Schneider (2013) prove that the payoff \( g_{sv} \) and \( g_{kns} \) can be replicated using European options. Following these works, we denote their (annualized) risk-neutral expectations by \( SVIX^2 \) and KNS,

\[
\frac{1}{T-t} \mathbb{E}_t^Q \left[ g_{sv}(r(t, T)) \right] = \frac{2}{T-t} \int_0^\infty \frac{\Theta_t(K, T)}{F_t(T)^2} dK =: SVIX_t^2 ,
\]

\[
\frac{1}{T-t} \mathbb{E}_t^Q \left[ g_{kns}(r(t, T)) \right] = \frac{6}{T-t} \int_0^\infty \frac{K - F_t}{K^2 F_t} \Theta_t(K, T) dK =: KNS_t.
\]

Contracts with payoff \( g_v \) of Equation (1) and \( g_{sv} \) of Equation (6) are both generalized variance contracts. The payoff functions differ only in the coefficients on terms higher than the second order. A linear combination of \( g_v \) and \( g_{sv} \) will cancel out the square of return,

\[
- \frac{3}{2} (g_v(r) - g_{sv}(r)) = - \frac{3}{2} \left( 2(e^r - 1 - r) - (e^r - 1)^2 \right) = r^3 + \frac{3}{4} r^4 + o(r^4).
\]

Similarly, the payoff function \( g_{kns} \) is also dominated by the cubed return. Applying Taylor expansion to \( g_{kns} \) given in Equation (6), we have

\[
g_{kns}(r) = r^3 + \frac{1}{2} r^4 + o(r^4).
\]

Notice that \( g_{kns}(r) \) and \( - \frac{3}{2} (g_v(r) - g_{sv}(r)) \) load differently only on the quartic and higher order terms of returns. Hence, we can easily separate the cubic returns from the quartic using linear transformation. The following proposition defines the payoff functions on the
quadratic, cubic, and quartic returns, and shows that they can be statically replicated using options. We denote the \( Q \)-expectation of the quadratic, cubic, and quartic returns by IM2, IM3, and IM4, respectively. The letter “I” is used to indicate their option-implied nature.

**Proposition 1.** Define \( g_2(r) \), \( g_3(r) \), \( g_4(r) \), as payoff functions of the second, third, and fourth power of returns, respectively:

\[
\begin{align*}
g_4(r) & := -6(g_v(r) - g_{sv}(r)) - 4g_{kns}(r) = r^4 + o(r^4), \\
g_3(r) & := g_{kns}(r) - \frac{1}{2}g_4(r) = r^3 + o(r^4), \\
g_2(r) & := g_v(r) - \frac{1}{3}g_3(r) - \frac{1}{12}g_4(r) = r^2 + o(r^4).
\end{align*}
\]

Under the condition of no arbitrage, the payoff \( \frac{1}{T-t}g_2(r(t,T)) \), \( \frac{1}{T-t}g_3(r(t,T)) \), and \( \frac{1}{T-t}g_4(r(t,T)) \) at time \( T \) can be statically replicated using a continuum of European options. The option-implied moments IM2\( _t \), IM3\( _t \), and IM4\( _t \) are defined as the time-\( t \) risk-neutral expectations of these payoffs:

\[
\begin{align*}
IM4\_t & := \frac{1}{T-t}E^Q_{t}\left[g_4(r(t,T))\right] = -6(VIX^2_t - SVIX^2_t) - 4KNS_t, \\
IM3\_t & := \frac{1}{T-t}E^Q_{t}\left[g_3(r(t,T))\right] = KNS_t - \frac{1}{2}IM4\_t, \\
IM2\_t & := \frac{1}{T-t}E^Q_{t}\left[g_2(r(t,T))\right] = VIX^2_t - \frac{1}{3}IM3\_t - \frac{1}{12}IM4\_t.
\end{align*}
\]

Different from \( g_v \) or \( g_{sv} \), which are exposed to higher order terms of returns, \( g_2 \) is a cleaner measure of quadratic return, since it does not contain the terms of cubic and quartic returns. Since both IM3 and IM4 are essentially linear combinations of VIX\(^2\), SVIX\(^2\) and KNS, they can be replicated using traded option prices in a model-free manner. In principal, we can use the Bakshi and Madan (2000) formula to include the fifth, sixth moment, etc. Similar to Schneider (2015), we truncate the higher moments at the fourth moment.

The realized counterparts of the aforementioned implied moments can be defined accordingly. Variables that start with “R” are meant to indicate their realized nature. We
define the realized simple variance (RSVₜ) based on simple returns and realized skewness measure (RKNSₜ) based on the insights in Kozhan, Neuberger, and Schneider (2013).

\[
\text{RSV}_t := \frac{1}{T-t} \sum_{i=1}^{n} (e^{r_i} - 1)^2, \\
\text{RKNS}_t := \frac{1}{T-t} \sum_{i=1}^{n} 6(2 + r_i - 2e^{r_i} + r_i e^{r_i}).
\] (11)

Let RM2, RM3, and RM4 denote the realized second, third, and fourth moment of returns, respectively. Based on RTV, RSV, and RKNS, the realized moments are given by

\[
\text{RM4}_t = -6(\text{RTV}_t - \text{RSV}_t) - 4\text{RKNS}_t, \\
\text{RM3}_t = \text{RKNS}_t - \frac{1}{2} \text{RM4}_t, \\
\text{RM2}_t = \text{RTV}_t - \frac{1}{3} \text{RM3}_t - \frac{1}{12} \text{RM4}_t.
\] (12)

Note that while RM2ₜ is not contaminated by the third and fourth moments, it still contains the moments which are higher than the fourth order. RM3ₜ and RM4ₜ also contain information about fifth and higher order moments. Taking the difference between the implied and realized moments gives the moment risk premium,

\[
\text{VRP}_t := \text{IM2}_t - \text{RM2}_t, \\
\text{M3RP}_t := \text{IM3}_t - \text{RM3}_t, \\
\text{M4RP}_t := \text{IM4}_t - \text{RM4}_t.
\] (13-15)

The following proposition decomposes the total variance risk premium, TVRP, into a second order or variance risk premium, VRP, and a higher order risk premium, HRP.

**Proposition 2** (Decomposition of TVRP). The total variance risk premium, TVRP, as given by Equation (5), can be decomposed into a variance risk premium VRP, and a
higher order risk premium $HRP$. That is,

$$TVRP_t = VRP_t + HRP_t,$$

where the higher order risk premium $HRP_t$ is defined as the sum of risk premiums on the third and fourth moments of returns,

$$HRP_t := \frac{1}{3} M3RP_t + \frac{1}{12} M4RP_t.$$  \hspace{1cm} (16)

The variance risk premium, $VRP$, is calculated as the difference between $TVRP$ and $HRP$. The higher order risk premium, $HRP$, is an aggregation of the moment risk premiums beyond the second order.$^3$ In this paper, we consider $HRP$ up to the fourth order.

3 Data Source and Risk Premium Measurement

3.1 Data Source and Variable Construction

We use the S&P 500 index option data from OptionMetrics, starting from January 1996 and ending in April 2016. Similar to the construction of VIX index provided by Chicago Board of Options Exchange (CBOE), we work with the best bid and ask closing quotes. Our data sample includes the highest closing bid and the lowest closing ask prices of all call and put options, strike prices, and expiration dates. We apply the standard filters to select the option sample. First, we delete all options with zero open interest, with zero bid prices, and with missing implied volatility. Second, following the literature on model-free implied volatility, such as Jiang and Tian (2005) and Carr and Wu (2009), we only keep out-of-the-money options. The definition of moneyness follows Martin (2017). A put option is out-of-the-money if the strike price is lower than the forward

$^3$We are aware that the rigorously defined skewness risk premium has a continuous trading term as in Kozhan, Neuberger, and Schneider (2013). However, we cannot implement the continuous trading term in our analytical decomposition framework.
price $F_t(T) = S_t e^{(r_f - q)(T - t)}$, where $S_t$ is the S&P 500 index spot price, $r_f$ is the risk free rate, and $q$ is the dividend rate. A call option is out-of-the-money if the strike price is higher than the forward price $F_t$. Third, we only keep options with less than 365 days of expiry. After applying the filters, we have 4,070,127 option-day data points. The option price is taken to be the average of the highest closing bid and lowest closing ask.

At the end of each month, we construct the annualized VIX$^2$, SVIX$^2$, and KNS using the discretized version of the formula discussed in Section 2 (see Equations (3), (7), and (8)). That is,

$$
VIX_t^2 \approx \frac{1}{T - t} \sum_{i=2}^{m} \left[ f(t, T, K_i) + f(t, T, K_{i-1}) \right] \Delta K_i,
$$

$$
SVIX_t^2 \approx \frac{1}{T - t} \sum_{i=2}^{m} \left[ f_s(t, T, K_i) + f_s(t, T, K_{i-1}) \right] \Delta K_i,
$$

$$
KNS_t \approx \frac{3}{T - t} \sum_{i=2}^{m} \left[ f_{kns}(t, T, K_i) + f_{kns}(t, T, K_{i-1}) \right] \Delta K_i,
$$

where $f(t, T, K_i) = \Theta_t(T, K_i) / K_i^2$, $f_s(t, T, K_i) = \Theta_t(T, K_i) / F_t^2$, $f_{kns}(t, T, K_i) = (K - F_t) \Theta_t(T, K_i) / K^2 F_t$, $\Delta K_i = K_i - K_{i-1}$, and $m$ is the total number of grid points in the strike dimension. $F_t$ denotes the forward price, and $\Theta_t(T, K)$ denotes the time $t$ value of an out-of-the-money option with strike price $K$ and maturity $T \geq t$. Following the construction of VIX provided by CBOE, we select two maturities of VIX$^2$, one with more than 30 days of expiry and one with less than 30 days and more than 7 days of expiry. Both of them have maturities closest to 30 days among all maturities. The annualized VIX$^2$ in Equation (17) is then calculated for these two maturities. Next, we interpolate the 30-day VIX$^2$ using VIX$^2$ of the two maturities with linear interpolation. The same procedure applies to the calculation of SVIX$^2$ and KNS with 30 days of expiration.

We use daily S&P 500 index prices from CRSP to estimate the realized total variance (RTV), realized simple variance (RSV), and realized KNS (RKNS). The three measures
are calculated using the following formula for each calendar month,

\[
RTV = 2 \sum_{i=1}^{N} (R_i - r_i), \quad RSV = \sum_{i=1}^{N} R_i^2, \quad RKNS = \sum_{i=1}^{N} 6(2r_i + (r_i - 2)R_i),
\]

where \( r_i = \log S_i - \log S_{i-1} - r_{f,i-1} \) is the excess daily log return, and \( R_i = \exp(r_i) - 1 \) is the simple return from the first trading day to the last trading day of each month. The simple return and log return on spot price in excess of the risk free rate are essentially futures returns, in line with Kozhan, Neuberger, and Schneider (2013).

### 3.2 Risk Premium Measurement

We calculate the total variance risk premium (TVRP) as the difference between \( \text{VIX}_2 \) and \( \text{RTV} \). \( \text{VIX}_2 \)—as the expected 30-day risk neutral variance—is calculated using OTM options at the last trading day of each month, and \( \text{RTV} \)—as the expected 30-day realized variance—is calculated as the sum of squared daily log returns within the month. We calculate the third and fourth return moments under the risk neutral measure (IM3 and IM4), the realized third and fourth return moments (RM3 and RM4), higher-order risk premiums (MRP3, MRP4, and HRP), and variance risk premium (VRP) in similar fashions.\(^4\) In Section 5.3, we discuss the difference between daily and high frequency data in constructing the higher order moments and their associated risk premiums.

Table 1 reports the summary statistics of option-implied moments (Panel A), realized moments (Panel B), and risk premiums (Panel C). Panel D reports the correlations between risk premiums. Comparing Panels A and B, we observe that the option-implied moments have larger means than their realized counterparts. \( \text{VIX}_2, \text{SVIX}_2, \) and IM2 are all generalized measures of option-implied variance that only differ in higher order moments of returns beyond the second order. These three measures share similar levels of mean, variance, skewness, and kurtosis, and so do their realized counterparts, RTV, RSV, and RM2. We also observe that, although realized variances are smaller on average than

\(^4\) IM2, IM3, and IM4 are calculated using Equation (10). RM2, RM3, and RM4 are calculated using Equation (12).
option-implied variances, they have a larger dispersion, in the sense that the standard deviation, skewness, and kurtosis are larger than their option-implied counterparts.

KNS and IM3 have leading exposures to the option-implied third moment of returns. In Section 6, we interpret higher order moments as reflecting rare events, or jumps, in asset returns. They all have negative mean and negative skewness, suggesting that investors expect more or larger negative jumps than positive jumps. Similar patterns can be observed in their realized counterparts, RKNS and RM3, suggesting similar asymmetry in realized jumps. Notice that compared with the difference between option-implied and realized second moments, the gaps between option-implied (IM3, IM4, and IHM) and realized (RM3, RM4, and RHM) higher moments are substantially larger. This indicates that rare events are clearly priced in the option prices.

Turning to Panel C, we can see that VRP and HRP have opposite signs. Consistent with the existing literature, VRP is on average positive with a mean of 1.12%. Since investors would hedge against the unfavorable negative jumps using options, the option-implied third moment IM3 is more negative than the realized third moment RM3. The risk premium associated with the third moment, M3RP, is on average negative with a mean of -0.68% in Panel C. The risk premium of the fourth moment, M4RP, has a positive mean of 0.24%. The intuition for positive M4RP is that investors consider the increase of variance-of-variance as deterioration of investment opportunity set (Merton, 1973). Closely related to M3RP, HRP is on average negative with mean -0.21%. Dominated by VRP, TVRP is positive but smaller than VRP.

Panel D of Table 1 calculates the correlations among risk premiums. The correlation between TVRP and VRP is as high as 0.99, implying that VRP is the major component of TVRP. It is not surprising that the correlation between M3RP and M4RP is also very high at -0.99, since the jump risk premium constitutes a large part of both.

Figure 1 plots the time series of TVRP, VRP, and HRP. The time series patterns of TVRP and VRP are almost indistinguishable. Both TVRP and VRP fluctuate between positive and negative values and display moderate variations as well as occasional spikes. Despite the fact that both TVRP and VRP are on average positive, as shown by the
summary statistics, there are a couple of extreme negative values in late 2002, 2008, and 2011. These negative spikes may be attributed to the downward volatility jumps as proposed by Amengual and Xiu (2018), associated with resolutions of policy uncertainties. Unlike TVRP or VRP, HRP is always negative and has less fluctuations. Moderate spikes in HRP coincide with large and volatile periods in VRP.

Figure 2 plots the autocorrelation function of VRP (Panel (a)) and HRP (Panel (b)), with the 95% confidence band indicated in the grey lines. Consistent with Figure 1, we observe that HRP is more serially correlated than VRP. The one-month autocorrelation coefficient for VRP is 0.4, while that for HRP is 0.6. The autocorrelation of VRP dies out after three month and becomes negligible in the medium and longer horizons. The autocorrelation of HRP, however, is significantly positive from 1- to 6-month horizon and remains positive (albeit not statistically significant) for as long as 15-month horizon.

4 Predictive Regression Analysis

In this section, we analyze the predictability of market returns by VRP and HRP as components of TVRP, both in-sample and out-of-sample. The superior performance of the joint VRP and HRP predictability largely remains if one excludes the global financial crisis period. Also, the joint predictability produces a significant utility gain in an asset allocation exercise. Moreover, the predicting power becomes even stronger for the established portfolio returns of the momentum factor.

4.1 Predicting Aggregate Market Return

As shown by Bollerslev, Tauchen, and Zhou (2009), Drechsler and Yaron (2010), and Bekaert and Hoerova (2014), TVRP has a significant in-sample predictive power for future market returns at quarterly horizon. This predictability tapers off for longer horizons beyond 6 months. We show that at any horizon from 1 month to 24 months, replacing TVRP with its VRP and HRP components yields better predictive results, both in sample and out of sample. Moreover, we find that VRP predicts short-term market returns up
to 9 months, while HRP predicts medium-term market returns up to 24 months. In multivariate predictive regressions, the coefficient of VRP is positive, and that on HRP is negative. This explains why TVRP alone, as the sum of the two components, has weaker predictive power over short horizons and hardly any over medium horizons.

We use the following specification for in-sample predictive regressions,

\[
r_{t,t+h} = \alpha_h + \beta_h'X_t + \varepsilon_{t,t+h},
\]

where \(r_{t,t+h}\) is the market excess return from the first day of next month \(t + 1\) to the last day of month \(t + h\), depending on the horizon \(h\). We use the excess log return on S&P 500 index as a proxy for market return. \(X_t\) is a vector of predictive variables, containing end-of-month values. In the univariate regressions, \(X_t = \text{TVRP}_t\) or \(\text{VRP}_t\) or \(\text{HRP}_t\). In the joint regression, \(X_t = (\text{VRP}_t, \text{HRP}_t)'\). We use Newey-West standard errors to correct for auto-correlation introduced by overlapping data.

While many variables can significantly predict market return in sample, most of them perform poorly in the out-of-sample (OOS) tests as shown in Welch and Goyal (2008). To assess the predictability of TVRP, VRP, and HRP more comprehensively, we also evaluate their OOS predictive performance. For each predictive regression, we calculate a return forecast as \(\hat{r}_{t,t+h} = \hat{\alpha}_h + \hat{\beta}_h'X_t\), where \(\hat{\alpha}\) and \(\hat{\beta}\) are OLS estimates from regression (18), using only the data available up to the time at which the forecast is made.

Following Welch and Goyal (2008), we define the OOS \(R^2\) for horizon \(h\) as

\[
\text{OOS } R^2_h = 1 - \frac{\text{MSE}_h}{\text{MSE}_{bm,h}}, \quad \text{and } \text{MSE}_h = \frac{1}{N} \sum_{i=1}^{N} (r_{t,t+h} - \hat{r}_{t,t+h})^2,
\]

where \(\text{MSE}_h\) is the mean squared forecast error of the predictive regression, and \(\text{MSE}_{bm,h}\) is the benchmark mean squared forecast error, which uses the average excess return from the beginning of the sample through month \(t\) as the return forecast for the next period. The OOS \(R^2\) measures the forecast accuracy relative to the benchmark. A positive OOS \(R^2\) implies out-performance relative to the naive forecast using the historical mean, and
a negative OOS $R^2$ implies under-performance.

The results of in-sample and out-of-sample forecasts are reported in Table 2, including univariate regressions using TVRP, VRP, and HRP, respectively, and multivariate regressions using VRP and HRP jointly. In the univariate regression of TVRP (first column of each horizon), the coefficient of TVRP is positive and highly significant for horizons of up to 6-month, with a maximum adjusted $R^2$ of 8% achieved at 3-month horizon. The coefficient of TVRP is not significant in horizons longer than 6-month. The OOS $R^2$’s are positive at 1- and 3-month horizons, but become negative for horizons beyond 3-month.

The predictive performance of VRP alone (second column of each horizon) is slightly better than that of TVRP at all horizons, in terms of higher $t$-statistics, higher adjusted $R^2$, and higher OOS $R^2$. At short horizons (1- and 3-month), VRP yields significant positive coefficients, substantial in-sample goodness-of-fit, and positive OOS $R^2$’s. At medium horizons, VRP is no longer significant and produces poor in sample and OOS $R^2$’s.

Interestingly, the predictive power of HRP (third column of each horizon) has a different pattern. The coefficients on HRP are invariably negative across all horizons. At the short end, the predictive regressions on HRP feature small $t$-statistics, low in-sample $R^2$, and negative OOS $R^2$. At medium horizons (9- to 24-month), by contrast, HRP exhibits stronger predictive power, with in-sample $R^2$ ranging from 2.92% to 4.38% and OOS $R^2$ ranging from 2.57% to 5.45%.

Turning to the multivariate regressions (fourth column of each horizon), the coefficients of VRP are always positive, and that on HRP are always negative. Different from the univariate regressions, where VRP is only significant at short horizons and HRP at long horizons, both coefficients are statistically significant at all horizons in the joint regression. Across all horizons, the adjusted $R^2$’s increase in the multivariate regressions compared with any univariate regressions or the combined $R^2$’s of univariate VRP and HRP regressions. Most notably, the OOS $R^2$’s are all positive at all horizons.
4.2 Further Discussions

Our evidence on univariate and multivariate predictive regressions underscores the importance of decomposing TVRP into HRP and VRP. First, the univariate regressions show that VRP contains short-term predictive power and HRP mostly medium-term predictive power. Since VRP and HRP predict future returns in opposite signs, the predictive power of TVRP is substantially hindered as a result of the negative prediction by HRP canceling out the positive prediction by VRP. This could explain why TVRP is not as a strong predictor as VRP at short horizons and has less predictive power at medium horizons than HRP.

Second, while the univariate regression results suggest that VRP has little predicting power at medium horizons and HRP at short horizons, the multivariate regressions suggest otherwise. At short horizons, the HRP coefficient turns statistically significant in the multivariate regression despite its insignificance in the univariate ones. For example, at the 3-month horizon, HRP and VRP jointly produce an in-sample $R^2$ as high as 14%, while the in-sample $R^2$ of VRP alone is only 9.55%.

Finally, separating VRP and HRP leads to significant improvement in out-of-sample forecasts. The OOS $R^2$'s of the joint regressions are positive across all horizons, while positive OOS $R^2$'s of TVRP are only found at 1- and 3-month horizons. Even at short horizons, where TVRP gives positive OOS $R^2$, VRP and HRP jointly deliver even higher OOS $R^2$ values. At 6-month horizon, in particular, the OOS $R^2$ of the multivariate regression reaches double digit 11%, in contrast to the negative values from univariate regressions.

To illustrate the improvement in prediction power across different horizons, Figure 3 plots the in-sample adjusted $R^2$ (top panel) and OOS $R^2$ (bottom panel), for the joint regressions with VRP and HRP (solid lines) versus the univariate regressions with TVRP (dashed lines), as functions of forecasting horizons. The joint regression beats the univariate regression in both in-sample adjusted $R^2$ and OOS $R^2$ across all horizons. For the horizons longer than 9-month, the in-sample $R^2$ of TVRP converges to 0, while that
of the joint regression remains above 5%. The OOS $R^2$ of TVRP is only positive at 1-, 3-, and 4-month horizons. For horizons longer than 4-month, the OOS $R^2$ of TVRP quickly deteriorates to negative values (around -10%). The OOS $R^2$'s of the joint regression, however, are positive at all horizons, with a maximum of 15% at quarterly horizon and approximately 5% at longer horizons up to 24 months.

The evidence indicates that identifying VRP and HRP as two factors helps to make more accurate OOS return forecasts, which could be exploited by investors. We further show that these forecasts lead to substantial utility gain for stock investors in Section 4.4.

4.3 Excluding Global Financial Crisis Period

To test the importance of crisis period in the predictability of HRP, in this section, we rerun the same predictive regressions after excluding the financial crisis period from December 2007 to June 2009. The estimated regression coefficients, $R^2$, adjusted $R^2$, and OOS $R^2$ are reported in Table 3. Results show that the predictive power of HRP in the univariate regressions is higher for 1 to 6-month horizons with higher $t$-statistics, higher adjusted $R^2$, and higher OOS $R^2$. However, the intermediate-horizon (9 to 24 months) predictability of HRP in both univariate and multivariate regressions is much weaker than in the full sample with insignificant $t$-statistics and lower adjusted $R^2$.

The sub-sample analysis excluding the global financial crisis period indicates the potential role of time-varying rare disaster probability in generating the medium-horizon predictability of HRP for market returns. In Section 6, we show how a model with rare event risk generates predictability over different horizons.

4.4 Asset Allocation

In this section, we measure the additional economic value of HRP’s predictive power beyond that of TVRP or VRP in an asset allocation perspective. Similar to Campbell and Thompson (2007), Rapach, Strauss, and Zhou (2010), and Rapach, Ringgenberg,
and Zhou (2016), we consider a mean-variance investor who allocates her wealth between a stock and a risk-free asset. At the end of month $t$, she invests in $w_t$ of her wealth in stocks and the rest in risk-free assets during the next month. The optimal weights are determined by forecasting the future stock returns using the predictive regressions:

$$w_t = \frac{\hat{r}_{t+h}}{\gamma \hat{\sigma}^2_{t+h}},$$

(19)

where $\gamma$ is the investor’s relative risk aversion coefficient, $\hat{r}_{t+h}$ and $\hat{\sigma}^2_{t+h}$ are the forecast of excess return and return variance $h$-month ahead. We use VIX$^2$ as a forward-looking measure of the excess return variance, because it reflects investors’ expectation of the return variation in the future. We consider two scenarios. In the first scenario with short-sale restriction, we impose a constraint for the portfolio weight to be larger than zero and smaller than 1.5. In the second scenario without short-sale restriction, similar to Campbell and Thompson (2007) and Rapach, Ringgenberg, and Zhou (2016), we restrict the portfolio weight $w_t$ to lie within the range of -0.5 to 1.5.

The certainty equivalent return (CER) for the investor who allocates assets using Equation (19) is: $\text{CER} = R_p - 0.5 \gamma \sigma_p^2$, where $R_p$ and $\sigma_p^2$ are the mean and variance of the (simple) portfolio return over the forecast evaluation period. Note that we focus on the simple excess return instead of the log excess return for the asset allocation analysis. We also compute the CER for the investor when she uses the historical mean as excess return forecast and VIX$^2$ as variance forecast. The CER gain is then the difference between the CER for the investor when she uses the predictive regression forecast and the CER when she uses the prevailing mean benchmark forecast. We annualize the CER gain so that it can be interpreted as the annual portfolio management fee that the investor would be willing to pay to have access to the predictive regression forecast in place of the prevailing mean forecast. In this way, we measure the direct economic value of return predictability.

To evaluate the CER gain of return predictability at longer horizons, we assume that the investor rebalances her portfolio weights at the same frequency as the forecast horizon. For the quarterly horizon, at the end of the quarter, the investor uses a predictive
regression or prevailing mean forecast of the excess return over the next three months and the allocation rule given by Equation (19) to determine the stock weight for the next quarter. The investor updates her portfolio weights at the end of next quarter, such that the return forecasts are nonoverlapping. For longer horizons from 6 to 12 months, the investor follows similar procedures to rebalance her portfolio.

The results of out-of-sample CER gains are reported in Table 4. We use the first half of the sample to estimate predictive regression parameters and the second half to evaluate out-of-sample portfolio performance. Similar to the calculation of OOS $R^2$, we employ an expanding window to estimate the predictive regression parameters and to calculate the historical average of excess return.

In Panel A, we report the CER gains of the investment strategy based on TVRP or VRP in the univariate predictive regression and HRP and VRP (denoted as HRP+VRP) in the bivariate predictive regression, when there is a short-selling constraint on the portfolio weight. All three predictive regressions are able to generate positive CER gains from 1-month to 9-month horizon with $\gamma = 3, 4$ and 5. It is worth noting that VRP almost always generate higher CER gains than TVRP except for the annual horizon, suggesting that the cleaner VRP measure has more economic value for a risk-averse investor than TVRP. HRP+VRP achieves the highest CER gain at all horizons with different levels of risk aversion, compared with univariate regressions and the buy-and-hold strategy.

In Panel B, we report the CER gains when the portfolio weights are updated without short-selling constraints. The result is similar to those in Panel A. Overall, the results in this section suggests that the information in TVRP, VRP, and the combination of HRP and VRP has substantial economic value for a risk-averse investor. VRP achieves higher performance than TVRP at horizons up to 6-month. The combination of HRP and VRP almost always generates the highest CER gains under different settings.

### 4.5 Predicting Returns of Factor Portfolios

We further explore whether VRP and HRP can predict returns of risk factors other than the aggregate market return. Table 5 reports the regression results for the size (SMB),
value (HML), and momentum (WML) portfolios for 3-month and 9-month horizons. Similar to the aggregate market portfolio, the 9-month predictability for the size portfolio comes from HRP exclusively. While the \( t \)-statistics associated with HRP are all significant at both horizons, the \( t \)-statistics for VRP are all insignificant. Moreover, the adjusted \( R^2 \)’s from the regressions that include both VRP and HRP are very close to those from the simple regressions with HRP only.

The results for the HML book-to-market portfolio reported in Table 5 show a different picture. HRP has strong predictive power for short-term return with \( t \)-statistics 5.36 at 3-month horizon in the univariate regression. The \( t \)-statistics of TVRP and VRP are all insignificant at all horizons in the univariate regressions. The adjusted \( R^2 \) appears to be maximized at the quarterly horizon at 9.00% for the joint regression with both VRP and HRP. The OOS \( R^2 \)’s, however, are almost always negative in all regressions.

We next turn to the WML momentum portfolio in Table 5. The coefficients of HRP are statistically significant for both horizons in the univariate regressions, with \( t \)-statistics 2.73 and 4.91. The coefficients of VRP, on the other hand, are significant only at 9-month horizon. While the sign of HRP is negative in the market return predictive regression, the sign of the HRP is positive in the return regression of WML. The adjusted \( R^2 \) of the bivariate regression is 8.21% at the quarterly horizon, with nearly all predictability attributable to HRP. The adjusted \( R^2 \)’s of the bivariate regression are more impressive at longer horizons: 24.11% at 9-month, with HRP the main contributor to predictability. It is noteworthy that the OOS \( R^2 \) are positive for HRP from 3- and 9-month horizons, in both univariate and bivariate regressions. Daniel and Moskowitz (2016) find that momentum strategies experience infrequent and persistent crashes, which typically occur following market declines. The high \( R^2 \) in the predictive regressions shows that rare events might offer potential explanation for the momentum anomaly, which we leave for future research.
5 Robustness Checks

In this section, we discuss a series of additional predictive regressions to address the robustness and validity of our empirical findings. We start by running predictive regressions with established predictors as control variables in Section 5.1. In Section 5.2, we study the predictability in the VRP and HRP constructed with predicted realized moments. Lastly in Section 5.3, we use 5-minute intraday returns to construct realized moments and test whether intraday moment premiums produce similar prediction results. Section 5.4 discusses additional robustness checks with technical details provided in Appendix B.

5.1 Control for Other Prediction Variables

To verify that the joint predictive power of VRP and HRP is robust after controlling for established return predictors, we consider a set of predictors documented in the previous literature as control variables. Specifically, 11 variables are considered: log dividend yield (DY), log earnings-price ratio (EP), book-to-market ratio (BM), the interest rate on a three-month Treasury bill (TBL), the difference between Moody’s BAA- and AAA-rated corporate bond yields (DFY), the long-term government bond yield (LTY), net equity expansion (NTIS), inflation calculated from the CPI for all urban consumers (INFL), the long-term government bond return (LTR), the long-term corporate bond return minus the long-term government bond return (DFR), and the log of the detrended short interest (SI).\(^5\) The first ten variables are discussed in Welch and Goyal (2008) and the last variable in Rapach, Ringgenberg, and Zhou (2016).\(^6\)

We report the results of return regressions for 3-month and 9-month horizons in Tables 6 with each of the 11 predictors as control variable in each column. In Panels A and B of Table 6, the coefficients of HRP and VRP are both statistically significant in all regressions. However, among the 11 control variables, only DY, BM and SI have

---

\(^5\)The construction of the short interest includes all publicly listed stocks on U.S. exchanges, ADRs, ETFs, and REITs, after excluding assets with a stock price below $5 per share and assets that are below the fifth percentile breakpoint of NYSE market capitalization.

\(^6\)The data of the predictor variables are obtained from Goyal and Rapach’s websites.
significant coefficients. The adjusted \( R^2 \)'s range from 14.75\% to 20.87\% in Panel A and from 9.57\% to 28.08\% in Panel B. Compared with the OOS \( R^2 \) of joint VRP and HRP regressions in Table 2 (14.52\%), the OOS \( R^2 \) increases only when TBL, LTY, NTIS or SI are added to the joint regressions. It is noteworthy that the OOS \( R^2 \) of joint VRP, HRP, and SI regression achieves 20.87\% at the quarterly horizon. In the 9-month regressions, combining VRP, HRP, and SI also results in improved in-sample and OOS performance, with adjusted \( R^2 \) 26.62\% and out-of-sample \( R^2 \) 29.92\%. In summary, after controlling for the 11 popular predictors, the joint predictability of VRP and HRP remains intact in sample and out of sample.

5.2 Predicted Realized Variances

The VRP measure used in Section 4.1 is calculated as the difference between the risk neutral variance at the end of the month and the realized variation of daily returns during that month. The advantage is that the measures are directly observable and model-free. However, by assuming that the expected realized variance in the next month is the realized variance in the past month, we implicitly assume the realized variance is a random walk. To be more consistent with the theoretical definition of VRP and HRP that TVRP\(_t\) = VIX\(^2\)_t - \( \mathbb{E}_t[RTV_{t+1}] \) and HRP\(_t\) = (VIX\(^2\)_t - IM2\(_t\)) - \( \mathbb{E}_t[RTV_{t+1} - RM2_{t+1}] \), we use forecasting models to get estimates of the conditional RTV in Equation (4) and conditional RM2 in Equation (12). Similar to Bekaert and Hoerova (2014), we consider the importance of persistence in realized variance using lagged realized variances as predictors and potentially different information in jump components such as lagged value of RM3. Our forecasting models can be represented as follows:

\[
RTV_t = \alpha_1 + \beta_1 RTV_{t-1} + \beta_2 RM3_{t-1} + \epsilon_t,
\]
\[
RM2_t = \alpha_2 + \beta_3 RM2_{t-1} + \beta_4 RM3_{t-1} + \epsilon_t.
\]

We use a moving window of 60 months to get one-step-ahead forecast of RTV and RM2 using regressions with their lagged values and RM3 as predictors. To obtain the
predicted RTV and RM2 from January 1996, we extend the data of RTV, RM2 and RM3 back to January 1991. Conditional expectation of RTV and RM2 can be estimated using

\[
E_t[RTV_{t+1}] = \hat{\alpha}_1 + \hat{\beta}_1 RTV_{t-1} + \hat{\beta}_2 RM3_{t-1},
\]

\[
E_t[RM2_{t+1}] = \hat{\alpha}_2 + \hat{\beta}_3 RM2_{t-1} + \hat{\beta}_4 RM3_{t-1}.
\]

where \(\hat{\alpha}_i, \hat{\beta}_j, i = 1, 2, j = 1, 2, 3, 4\), are OLS estimators. We use a superscript \(f\) to denote quantities that are constructed using the forecast of realized moments of Equation (20).

Table 7 reports the estimated regression coefficients, \(R^2\), adjusted \(R^2\), and OOS \(R^2\) of the predictability regressions from one to 24-month excess returns on the S&P 500 index. For each horizon, we report predictive regression results of different predictors: TVRP\(^f\), VRP\(^f\), HRP\(^f\), and VRP\(^f\) and HRP\(^f\) jointly. TVRP\(^f\) is calculated as the difference between VIX\(^2\) and the predicted value of RTV. VRP\(^f\) is the difference between IM2 and the predicted value of RM2. HRP\(^f\) is the difference between TVRP\(^f\) and VRP\(^f\).

Although the predictability in Table 7 is in general weaker than the results using “model-free” realized variances, the coefficients of HRP\(^f\) in univariate and multivariate regressions are all significantly negative from 6-month horizon to 24-month horizon. Decomposing TVRP\(^f\) into VRP\(^f\) and HRP\(^f\) improves adjusted \(R^2\) and out-of-sample \(R^2\) from 6-month horizon to 24-month horizon. Overall, we find that decomposing TVRP calculated based on forecasted variance improves predictive performance both in-sample and out-of-sample in intermediate horizon.

### 5.3 High Frequency Realized Measures

As shown in Neuberger and Payne (2017), skewness estimates from moving windows of daily or weekly data are likely to have different averages than those constructed from intraday data due to leverage effect. This is in contrast to the second moment, where using higher frequency returns yields more efficient estimates of quadratic variation (see, e.g., Andersen, Bollerslev, Diebold, and Ebens (2001), Barndorff-Nielsen and Shephard (2002)). In this section, we use 5-minute intraday returns to construct realized moments.
and test whether the moment premiums constructed from high frequency data can produce similar prediction results, compared to those constructed from daily data in the previous sections.

The realized total variance (RTV\textsuperscript{5-min}), realized simple return variance (RSV\textsuperscript{5-min}), and realized KNS (RKNS\textsuperscript{5-min}) are calculated from the 5-min intraday returns. Similar to the decomposition in Section 2.2, the intraday total variance risk premium, TVRP\textsuperscript{5-min} is decomposed into VRP\textsuperscript{5-min} and HRP\textsuperscript{5-min}. We report the summary statistics of the high frequency measures in Table 8 and the predictive regression results in Table 9.

Consistent with Amaya, Christoffersen, Jacobs, and Vasquez (2015) and Neuberger and Payne (2017), we observe that the realized third and fourth moments have different means for daily and intraday data. Table 8 Panel A shows that the mean of the intraday RM3\textsuperscript{5-min} is $-1.5 \times 10^{-5}$, while the mean of the daily RM3 is $-1 \times 10^{-4}$ in Table 1. The mean of the intraday RM4\textsuperscript{5-min} is $2.83 \times 10^{-6}$, while the mean of the daily RM4 is $1 \times 10^{-4}$. Table 8 Panel B shows that the correlation between the daily and intraday second moment is as high as 0.96, while the correlation between the daily and intraday third moment is only -0.03. This evidence further confirms that the realized third moments constructed from high and low frequency data contain distinct information.

Table 9 reports the predictive regression results using realized moments from intraday data. We see that TVRP\textsuperscript{5-min} and VRP\textsuperscript{5-min} outperform their daily counterparts in univariate regressions, which is consistent with the evidence in Bollerslev, Tauchen, and Zhou (2009). The predictive performance of HRP, by contrast, deteriorates dramatically when using intraday data (Table 9), compared to using daily data (Table 2), with smaller $t$-statistics, lower in-sample $R^2$’s, and negative OOS $R^2$’s.

We argue in Section 6.2.4 that the reason of the weak predictive power in intraday HRP lies in the fact that it is a less prominent measure of the rare consumption disaster intensity. High-frequency third moment only captures the third moment of jumps in the stock return (Amaya, Christoffersen, Jacobs, and Vasquez, 2015), while the option-implied third moment also contains the covariance between return and variance innovations (the “leverage effect”). Since the high-frequency realized third moment fails to
match the leverage term in option-implied third moment, HRP\textsuperscript{5-min} is a less effective predictor than its daily counterpart.

5.4 Additional Robustness Checks

We next implement two alternative methods to decompose TVRP. First, we decompose TVRP with the cubic and quartic contracts defined in Bakshi, Kapadia, and Madan (2003) and find similar results (Appendix B.1). Second, we decompose TVRP using the moment risk premium calculated from the swaps introduced by Schneider and Trojani (2015) and Schneider and Trojani (2017). These swaps have the Aggregation Property, meaning that the floating legs can be dynamically replicated in a way that is robust to the sampling frequency of returns. We show that this decomposition method also yields qualitatively similar results, implying that the aggregation property does not make much difference in terms of forecasting performance (Appendix B.2). The contracts introduced by Schneider and Trojani (2015) also provide a more generalized approach to decompose TVRP into an arbitrary order of moment risk premium.

Finally, we show that VRP and HRP jointly give better forecasting performance than ad hoc jump- or skewness-related predictors proposed in the literature. See Appendix B.3 for the alternative candidate predictors and their respective predictive performance.

6 Structural Interpretation with Rare Events

To explore the underlying economic mechanism for explaining the predictability offered by HRP, which is different from yet complementary to that offered by VRP, we extend the consumption-based model by Bollerslev, Tauchen, and Zhou (2009) with stochastic volatility-of-volatility to allow for time-varying rare events.

We adopt the rare jump specification as in Wachter (2013), where the stochastic jump intensity follows a square root process. Jumps in our model have the interpretation of rare events in consumption risk, as in Liu, Pan, and Wang (2004) and Barro (2006).

In this section, we first explain the model setup and the solution technique. Then,
we analyze the economic variables of interest implied by the model, including VRP and HRP. Lastly, we study the model-implied predictability patterns of VRP and HRP and discuss the determinants of their differential predictive powers over different horizons.

6.1 Model Setup

The underlying environment is a discrete time endowment economy. The representative agent’s preferences on the consumption stream are of the Epstein and Zin (1989) form, allowing for the separation of risk aversion \( \gamma \) and the intertemporal elasticity of substitution \( \psi \). The log of the pricing kernel can be expressed as

\[
\begin{align*}
    m_{t+1} &= \theta \log \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1)r_{t+1}.
\end{align*}
\]

Here, \( 0 < \delta < 1 \) reflects the agent’s time preference, \( \theta = \frac{1 - \gamma}{1 - \psi} \), \( \Delta c_{t+1} \) is log consumption growth, and \( r_{t+1} \) is the log return of the consumption claim.

The dynamics of log consumption growth follows

\[
\begin{align*}
    \Delta c_{t+1} &= \mu_c + \rho_c \Delta c_t + \sigma_c Z_{c,t+1} + J_{t+1} \Delta N_{t+1}, \\
    \sigma_{t+1}^2 &= \mu_\sigma + \rho_\sigma \sigma_t^2 + \sqrt{\mu_t} Z_{\sigma,t+1}, \\
    q_{t+1} &= \mu_q + \rho_q q_t + \phi_q \sqrt{q_t} Z_{q,t+1}, \\
    \lambda_{t+1} &= \mu_\lambda + \rho_\lambda \lambda_t + \phi_\lambda \sqrt{\lambda_t} Z_{\lambda,t+1}.
\end{align*}
\]

Here, \( \sigma_t^2 \) and \( q_t \) are the two volatility factors along the lines of Bollerslev, Tauchen, and Zhou (2009). \( Z_{i,t+1} \sim N(0, 1), i = c, \sigma, q, \lambda \), are independent Gaussian shocks, \( N_{t+1} \) is a Poisson process with time-varying intensity \( \lambda_{t+1} \), and \( J_{t+1} \) is an i.i.d random variable representing consumption jump amplitude with a fixed probability density \( \nu \). We use the notation \( \mathbb{E}_\nu[f(j)] \) to denote the jump transform \( \int f(j)\nu(j)dj \).

Rare events in consumption are captured by the Poisson process \( N_{t+1} \), allowing for large instantaneous changes in \( c_{t+1} \). \( \lambda_t \) is the probability of a rare event over the period of \([t, t+1]\). Should a rare event occur, the log consumption \( c_{t+1} \) changes by \( J_{t+1} \). Because
rare events usually have a negative skew, we suppose that $J_{t+1}$ follows a negative Gamma distribution $J_{t+1} \sim -\Gamma(\nu_j, \frac{\mu_j}{\nu_j}) + a$. Similar to Liu et al. (2004), we focus our attention on undesirable event risk and assume $E[\nu] < 0$.

Our model is related to Drechsler and Yaron (2010) and Guo, Sha, Wang, and Zhou (2018). Both of these models introduce stochastic jumps in the long-run risk model by Bansal and Yaron (2004). We deviate from these models in two aspects—we do not include the long run risk, and we do not specify the jump intensity as proportional to the instantaneous variance of consumption growth. The long run risk, although important for many economic phenomena, is not essential for generating the medium-horizon predictability in HRP. Instead, we adopt a more parsimonious persistent consumption process as in Equation (22), to introduce the dependence between future rare event strikes and current rare event probability, which is critical to generate the medium term predictability of HRP for market returns.\footnote{Similar specification can be found in Bansal and Lundblad (2002), who model the dividend growth (which is also the consumption growth in our model) process as a stationary ARIMA process, in which the dividend (consumption) growth rate contains a predictable and persistent component.} Moreover, we model the event intensity as a separate process in order to isolate the effect of rare events from that of stochastic consumption volatility. In the special case that $Z_\lambda$ is the same as $Z_\sigma$, the event intensity $\lambda_t$ is perfectly correlated with $\sigma_t^2$, as in Drechsler and Yaron (2010) and Guo, Sha, Wang, and Zhou (2018). While including a long run risk component and perfectly correlated consumption variance and event intensity should be easy to incorporate in our framework, our specification (22) is not only more parsimonious but also critical to drive the differential information contents of VRP and HRP.

We collect the consumption growth $\Delta c_t$ and the latent state variables in the state vector $Y_t = (\Delta c_t, \sigma_t^2, q_t, \lambda_t)$, which follows

$$Y_{t+1} = \mu + FY_t + G_t Z_{t+1} + e_c J_{t+1} \Delta N_{t+1},$$

where $\mu = (\mu_c, \mu_\sigma, \mu_q, \mu_\lambda)'$, $F$ a constant diagonal matrix with $(\rho_c, \rho_\sigma, \rho_q, \rho_\lambda)$ on the diagonal, $G_t$ a diagonal matrix with $(\sigma_t, \sqrt{q_t}, \sqrt{q_t}, \sqrt{\lambda_t})$ on the diagonal, and $e_c = \ddots$
a selection vector that selects the consumption element.

Similar to Bollerslev, Tauchen, and Zhou (2009), we assume that the dividend growth rate is identical to the consumption growth rate. We solve the model using the standard log-linear approximation proposed by Campbell and Shiller (1988), which allows us to express the log return of equity as a linear function of the price-dividend ratio $w_t$:

$$r_{t+1} = \kappa_0 + \kappa_1 w_{t+1} - w_t + \Delta c_{t+1}. \quad (23)$$

The standard solution method for finding the equilibrium then consists of conjecturing a solution for $w_t$ as an affine function of the state variables $w_t = A_0 + A_1' Y_t$, where $A_1 = (A_c, A_\sigma, A_q, A_\lambda)'$ is a vector of pricing coefficients. We will solve for $A_0, A_1$ using the Euler equation

$$E_t[\exp(m_{t+1} + r_{t+1})] = 1.$$ 

Appendix C provides more details of the derivation.

### 6.2 Model Implication

#### 6.2.1 Pricing Kernel

The innovation to the pricing kernel is given by

$$m_{t+1} - E_t[m_{t+1}] = -\Lambda' \left( G_t Z_{t+1} + (J_{t+1} \Delta N_{t+1} - \lambda_t E_\nu[J_{t+1}]) \right). \quad (24)$$

Here, $\Lambda = (\Lambda_c, \Lambda_\sigma, \Lambda_q, \Lambda_\lambda)' = (1 - \theta)\kappa_1 A_1 + \gamma e_c$, can be interpreted as the price of risk for Gaussian shocks and also the sensitivity of the intertemporal marginal rate of substitution to the rare events. According to the Girsanov Theorem, $\tilde{Z}_t \equiv Z_t + G'_t \Lambda$ is a martingale under the risk neutral $Q$. Under the risk neutral measure, the counting process $N_{t+1}$ has intensity $\tilde{\lambda}_t = \lambda_t E_\nu[\exp(-\Lambda_c J_{t+1})]$, and the jump amplitude follows distribution $\tilde{\nu}(j) = \frac{\exp(-\Lambda_c j)}{\lambda_0} \nu(j)$. In case of the Gamma distribution, the jump distribution under $Q$ is given by $J \sim -\Gamma(\nu_j, \frac{n_j}{\nu_j - \Lambda_c j}) + a$. For a positive price of consumption risk $\Lambda_c$, negative
jumps are assigned a larger probability under $Q$. In addition, if consumption rare events are on average disastrous, then the event intensity multiplier $E\nu[\exp(-\Lambda c J_{t+1})]$ is larger than 1. Under usual parameterization, disastrous consumption events are more probable and happen more frequently under the pricing measure $Q$.

### 6.2.2 Equity Risk Premium

The innovation to the market return satisfies

$$r_{t+1} - \mathbb{E}_t[r_{t+1}] = B'G_tZ_{t+1} + (J_{r,t+1}\Delta N_{t+1} - \lambda_t E\nu[J_{r,t+1}]),$$

where $B = \kappa_1 A_1 + \epsilon_c$ and $J_{r,t+1} = (\kappa_1 A_1 + 1)J_{t+1}$ is the jump amplitude of market return. Since jump sizes are independent, we will drop the $t$ subscript of jump sizes when no confusion is caused. Let $R_{t+1}$ be the gross return on the market, following Drechsler and Yaron (2010), the conditional equity risk premium can be expressed as

$$\log(\mathbb{E}_t[R_{t+1}]) - r_{f,t} = B'G_tG_t'\Lambda + \lambda_t(\Psi(\epsilon_c'B) - 1) - \lambda_t(\Psi(\epsilon_c(B - \Lambda)) - \Psi(-\epsilon_c'\Lambda)) =: c_{\sigma}^{ERP} \sigma_t^2 + c_q^{ERP} q_t + c_{\lambda}^{ERP} \lambda_t + c_j^{ERP} \lambda_t,$$

where $\Psi(\cdot)$ is the moment generating function of the jump size given by $\Psi(t) \equiv \mathbb{E}_\nu[\exp(tJ)] = (1 + \frac{\mu_j}{\nu_j} t)^{-\nu_j} \exp(a)$. See Appendix C for explicit expressions of $c_{\sigma}^{ERP}$, $c_q^{ERP}$, $c_{\lambda}^{ERP}$, and $c_j^{ERP}$.

The first two terms, $c_{\sigma}^{ERP} \sigma_t^2$ and $c_q^{ERP} q_t$, represent the equity risk premium of the Gaussian shocks. They originate from the covariance between the innovation of the pricing kernel and that of the market return. Since stochastic volatility and volatility-of-volatility are priced risks in the pricing kernel, they both contribute to the equity risk premium. The third term, $c_{\lambda}^{ERP} \lambda_t$, represents the risk premium of time-varying jump intensity. Should rare events occur with a fixed probability, $A^2\phi^2_{\lambda}$ would be zero, in which case this term becomes trivial. The reason that it is proportional to $\lambda_t$ is that intensity follows a square-root process, i.e., the variance of the intensity is proportional to intensity itself. The last term, $c_j^{ERP} \lambda_t$, represents the risk premium of rare events. It is determined
by instantaneous consumption decline in case of a rare event and the probability of such events.

Under usual parametrization, where $\gamma > 1$ and $\psi > 1$, $c^{ERP}_\sigma$, $c^{ERP}_q$, $c^{ERP}_\lambda$ are all positive. The sign of $c^{ERP}_j$ depends on the sign of the expected jump amplitude. During turbulent periods, when state variables $\sigma_t, q_t, \lambda_t$ take high values, investors are compensated accordingly.

### 6.2.3 Variance Risk Premium

Next we derive the variance risk premium (VRP). Let $\hat{r}_{t+1}$ be the innovation of market return, $\hat{r}_{t+1} \equiv r_{t+1} - \mathbb{E}_t[r_{t+1}]$. If the unit time interval is regarded as half a month, we can define the variance risk premium as

$$
VRP_t \equiv \text{Var}_t(\hat{r}_{t+1} + \hat{r}_{t+2}) - \text{Var}_t(\hat{r}_{t+1} + \hat{r}_{t+2}),
$$

where $\text{Var}_t(\hat{r}_{t+1} + \hat{r}_{t+2})$ is the ex ante variance of the return next month. In Appendix C.4 we show that under the definition of Equation (26), VRP has both the level difference and drift difference components, as in Drechsler and Yaron (2010). More importantly, it allows for multiple returns generated within one month, thus enabling interaction between $\hat{r}_{t+1}$ and $\hat{r}_{t+2}$, which becomes essential when calculating the higher order moments. See the next subsection for details.

One can show that VRP can be written as

$$
VRP = c^{vrp} + c^{vrp}_q q_t + c^{vrp}_\lambda \lambda_t,
$$

with $c^{vrp}, c^{vrp}_q, c^{vrp}_\lambda$ explicitly expressed in Appendix C.

### 6.2.4 Higher Order Risk Premium

In this section, we derive the higher order risk premium under the special case of the two-period-per-month setting. We leave the more general case to Appendix C.7. When
there are two period returns in a month, the third moment of the monthly log return is defined as

\[ M_3_t \equiv \mathbb{E}_t[(\hat{r}_{t+1} + \hat{r}_{t+2})^3] \]

\[ = \mathbb{E}_t[\hat{r}_{t+1}^3] + \mathbb{E}_t[\hat{r}_{t+2}^3] + 3\mathbb{E}_t[\mathbb{E}_{t+1}[\hat{r}_{t+1}^2\hat{r}_{t+2} + \hat{r}_{t+1}\hat{r}_{t+2}^2]] \]

\[ = \mathbb{E}_t[\hat{r}_{t+1}^3] + \mathbb{E}_t[\hat{r}_{t+2}^3] + 3\text{Cov}_t[\hat{r}_{t+1}, \text{Var}_{t+1}[\hat{r}_{t+2}]]. \]

The last equality holds since \( \mathbb{E}_t[\hat{r}_{t+2}^3] = 0 \). As \( \mathbb{E}_t[\hat{r}_{t+1}^3] = \lambda_t \mathbb{E}_\nu[(\tilde{J}_r - \mathbb{E}_\nu[\tilde{J}_r])^3] \), the first two terms are only concerned with rare events. The third term is a leverage term—it is related to the covariance between the log return and return variance. The leverage term only appears when there are multiple returns generated in a month. That explains why we set the unit time interval to be half-a-month.

\[ \text{Cov}_t[\hat{r}_{t+1}, \text{Var}_{t+1}[\hat{r}_{t+2}]] \]

\[ = \kappa_1 \left( A_y B' H_y B \text{Var}_t(\sigma_{t+1}^2) + A_y B' H_y B \text{Var}_t(q_{t+1}) + (A_{1} H_{1} B + \text{Var}(J_r))\text{Var}_t(\lambda_{t+1}) \right) \]

(28)

The leverage term given by Equation (28) is under the physical measure \( P \). The \( Q \)-leverage term only differs from Equation (28) in the last term iii. The reason for this is that \( \sigma_t^2, q_t \) are only driven by Gaussian shocks and therefore have the same variance under \( P \) and \( Q \). As a result, the leverage premium (i.e., difference between \( Q \)- and \( P \)-leverage) is only a function of \( \lambda_t \), but not of \( \sigma_t \) or \( q_t \).

The risk premium associated with the third order, and approximately higher order risk premium (HRP), can be computed as

\[ \text{HRP}_t = \frac{1}{3} \left( \mathbb{E}_t^Q[M_{3, t+1}^Q] - \mathbb{E}_t[M_{3, t+1}] \right) \]

\[ = \frac{1}{3} \left( (\mathbb{E}_t^Q[\hat{r}_{t+1}^3] - \mathbb{E}_t[\hat{r}_{t+1}^3]) + (\mathbb{E}_t^Q[\hat{r}_{t+2}^3] - \mathbb{E}_t[\hat{r}_{t+2}^3]) \right) \]

\[ = c_{\text{hrp}} + \lambda_t \text{hrp}, \]  

(29)
where \( c_{\text{hrp}}, c_{\lambda_{\text{hrp}}} \) are constants given in Appendix C.

Note that the two-period-per-month HRP given by Equation (29) is a special case in which the volatility-of-volatility is not present in HRP. In the more general case where there are more than two periods in a month, HRP is also loaded on \( q_t \). However, the HRP’s loading on rare event intensity \( \lambda_t \) is still more prominent than its loading on volatility-of-volatility \( q_t \). See Appendix C.7 for details.

Although the leverage effect makes up a substantial fraction of the third moment, by taking the differences of \( Q \)- and \( P \)-higher moments, a great deal of the exposure to the volatility-of-volatility is mitigated. In the special case of two periods per month presented in this section, \( q_t \) is even completely canceled out in HRP, as shown by Equation (29).

This argument also explains why HRP calculated using intraday data underperforms HRP calculated using daily data. As the \( P \)-leverage term is absent in intraday return higher moments, the \( Q \)-leverage term is not taken away as much when subtracting the \( P \)-moments. Therefore, HRP given by intraday data is a less prominent measure of \( \lambda_t \).

### 6.3 Return Predictability from VRP and HRP

In this section, we consider the predictive regression for market returns,

\[
\sum_{j=1}^{2h} r_{t+j} = \beta_0 + \beta_1 X_t + \epsilon_{t+j,t},
\]

where \( h \) is the forecasting horizon measured in months, \( X_t \) the predictor, \( \beta_0 \) the intercept, and \( \beta_1 \) the predictive coefficient. The predictive coefficient \( \beta_1 \) is given by,

\[
\beta_1(h) = \frac{\sum_{j=1}^{2h} \text{Cov}[r_{t+j}, X_t]}{\text{Var}(X_t)}. \tag{30}
\]

The coefficient of determination is calculated as

\[
R^2(h) = \frac{\left( \sum_{j=1}^{2h} \text{Cov}[r_{t+j}, X_t] \right)^2}{\text{Var}(\sum_{j=1}^{2h} r_{t+j})\text{Var}(X_t)}.
\]
Since VRP and HRP are linear functions of $q_t$ and $\lambda_t$, predictability in VRP and HRP boils down to predictability in $q_t$ and $\lambda_t$. The covariance term in Equation (30) can be expanded into three parts (in the brackets),

$$
\text{Cov}[r_{t+j}, X_t] = (B'F - A_1') \left( F^{j-1} \text{Cov}[Y_t, X_t] + \sum_{i=1}^{j} F^{j-i} G_{t+i-1} \text{Cov}[Z_{t+i}, X_t] + \sum_{i=1}^{j} F^{j-i} e_c \text{Cov}(J\Delta N_{t+i}, X_t) \right).
$$

(31)

The first term is the covariation of the predictor $X_t$ with the state variable $Y_t$. When $q_t$ and $\lambda_t$ are used as predictors, this term is proportional to the persistence level of the predictors. The second term represents the covariation of the predictor with future Gaussian shocks and is zero by assumption. The third term captures the covariation of the predictor with future rare events. When the volatility-of-volatility is used as the predictor, i.e., $X_t = q_t$ in Equation (31), this term is trivial,

$$
\text{Cov}(J\Delta N_{t+i}, q_t) = 0.
$$

The predictability in $q_t$ thus comes from the first term on the RHS of Equation (31) only.

An important property of rare events is that the strike of a rare event in the future is correlated with the event probability at present. When the event intensity is used as the predictor, i.e., $X_t = \lambda_t$ in Equation (31), the $i^{th}$ summation element in the last term of the RHS of Equation (31) can be expressed as

$$
(B'F - A_1') F^{j-i} e_c \text{Cov}(J\Delta N_{t+i}, \lambda_t) = (B'F - A_1') F^{j-i} e_c \mathbb{E}[\text{Cov}(t+i-1(J\Delta N_{t+i}, \lambda_t)] + \mathbb{Cov}(t+i-1(J\Delta N_{t+i}, \lambda_t)) = \left((\kappa_1 \rho_c - 1) A_c + \rho_c \right) \rho_i^{j-i} \rho_{\lambda}^{i-1} \mathbb{E}[J] \text{Var}(\lambda_t).
$$

(32)

Observe that Equation (32) is not only a function of the persistence of event intensity, $\rho_{\lambda}$, but also of the persistence of consumption growth, $\rho_c$. As a matter of fact, if we
replace the expression for $A_c$. Equation (32) can be written as $\frac{1}{2} \rho^c \rho^{-1} \mathbb{E}_v[J] \text{Var}(\lambda_t)$, which takes a negative value given negative average jump size $\mathbb{E}_v[J]$. The intuition is that higher current event intensity anticipates more probable rare events in the future. Since these events are most often disastrous, future market returns are likely to decline.

We conclude that the predictability in $\lambda_t$ consists of two components. The first component is its own persistence. Probability of rare events at present, $\lambda_t$, is correlated with the rare event probability in the future, $\lambda_{t+j}$. This component is always positive, implying that higher current event intensity leads to larger equity premium in the future. The second component comes from the covariance between the current event intensity, $\lambda_t$, and future realization of rare events $\Delta N_{t+j}$. This term is negative, implying that higher current event intensity could cause future market declines. The eventual sign of $\text{Cov}(\lambda_{t+j}, \lambda_t)$ depends on which component is stronger at horizon $j$. The predictability in $q_t$, by contrast, comes from its own persistence alone. This crucial difference between $q_t$ and $\lambda_t$ drives their predictive powers over different horizons.

Calibrating the parameters to match the empirical predictive coefficients and $R^2$ is challenging in a high dimensional model like this and is beyond the scope of this paper. We shall illustrate that our model is capable of qualitatively reproducing the empirical patterns of the predictive coefficients and $R^2$ for VRP, HRP, and TVRP over different forecasting horizons. Table 10 reports parameter values used to produce Figure 4, Figure 5, and Figure 6. These baseline parameters are set to typical values used in Bollerslev, Tauchen, and Zhou (2009), Drechsler and Yaron (2010), Guo, Sha, Wang, and Zhou (2018), and Wachter (2013). Parameters should be interpreted as half-a-month quantities. Of particular interest is that, the volatility-of-volatility process $q_t$ is set to a moderate persistence level $\rho_q = 0.7$, while the jump intensity process $\lambda_t$ to a high persistence level $\rho_\lambda = 0.95$. These baseline persistent values help deliver the observed patterns in the empirical regressions. Appendix C.6 provides details on how we calibrate the parameters.

Figure 4 plots the model-implied covariance of return $h$-month away, $r_{t+2h}$, with $q_t$ (left panel) and $\lambda_t$ (right panel). We plot the covariance as a function of horizon $h$ (measured in months) for different persistence levels. On the left panel, we consider $\rho_q =$
0.3 (the dotted line), 0.5 (the solid line), and 0.7 (the dashed line). On the right panel, we consider $\rho_\lambda = 0.6$ (the dotted line), 0.9 (the solid line), and 0.95 (the dashed line). The covariance of $q_t$ and $\lambda_t$ with future returns are both positive at short horizons. The persistence parameters control the rate at which the covariances decrease with horizon. Note that the covariance between $\lambda_t$ and future returns can get negative for some values of $\rho_\lambda$ at certain horizons.

Next, we turn to predictive power of VRP and HRP. Under our parametrization, VRP’s loading on $\lambda_t$, $c^{\text{vrp}}_\lambda$, takes the value of 5.5, while its loading on $q_t$, $c^{\text{vrp}}_q$, is as large as $3.1 \times 10^5$. We therefore conclude that VRP is dominated by $q_t$. In the special case of two-period-per-month as in Section 6.2.4, HRP is a clean reflection of the event intensity $\lambda_t$, in which $c^{\text{hrp}}_q = 0$ and $c^{\text{hrp}}_\lambda = -0.52$. Even in the more general cases with more than two periods in a month in which case HRP is a linear function of both $q_t$ and $\lambda_t$, HRP is still a more prominent measure for $\lambda_t$ than VRP (see Appendix C.7 for details). Therefore VRP mostly exhibits the short-term predictability in $q_t$, whereas HRP displays a longer, medium-horizon predictability in $\lambda_t$.

Figure 5 plots the model-implied predictive coefficient ($\beta_1$ in Equation (30)) on VRP (left panel) and HRP (right panel) as a function of horizon $h$ (measured in months) for the same persistence levels as in Figure 4. As expected, in both panels, larger persistence leads to larger magnitude of the predictive coefficients. The VRP beta is always increasing with horizon whatever the persistence of $q_t$ is. Interestingly, the HRP beta displays different shapes under different persistence levels of $\lambda_t$. In the empirical regressions, the HRP betas are negative and decrease with horizon, implying high persistence level of the event intensity. This is consistent with the literature on rare disasters—event intensity is typically found to be persistent (Wachter, 2013). More importantly, the predictive coefficients on VRP increase mildly from 1- to 24-month, while the magnitude of HRP betas has increased sharply. This pattern is also consistent with our empirical finding.

Figure 6 plots the model-implied $R^2$ as a function of predictive horizons (measured in months) for TVRP alone (dotted line) and VRP and HRP jointly (solid line). Similar to the empirical $R^2$ depicted in Figure 3, the model-implied $R^2$’s of both regressions
are hump-shaped, peaking at 6-month. Moreover, the model-implied $R^2$ of the joint regression stays above that of TVRP for all horizons.

7 Conclusion

The conventional variance risk premium has been shown to be a strong predictor for market returns in recent literature. We provide a model-free, internally consistent decomposition of the total variance risk premium (TVRP) into a second order (variance) risk premium (VRP) and a higher order risk premium (HRP).

VRP (HRP) contains short-term (medium-term) predictive power for market returns with statistically significant positive (negative) coefficients. When used jointly as predictors, VRP and HRP deliver higher in-sample and out-of-sample $R^2$'s than TVRP alone, across all horizons from 1- to 24-month. This result survives the exclusion of the global financial crisis period, generates significant economic gain in an asset allocation exercise, and proves to be even stronger for the portfolio returns of the momentum factor.

We rationalize these empirical findings in a Lucas type equilibrium economy with stochastic volatility-of-volatility and time-varying consumption rare events. Such a framework endogenously generates positive and negative predictive coefficients on VRP and HRP, respectively. The model implied hump-shaped $R^2$'s also match the empirical ones for the joint VRP and HRP regressions. The short term predictability in VRP can be explained by the volatility-of-volatility process with moderate persistence, while more persistent rare event intensity contributes to the medium-horizon predictability in HRP.
References

D. Amaya, P. Christoffersen, K. Jacobs, and A. Vasquez. Does realized skewness predict


T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized


G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws, and the
differential pricing of individual equity options. *Review of Financial Studies*, 16(1):

R. Bansal and C. Lundblad. Market efficiency, asset returns, and the size of the risk

R. Bansal and A. Yaron. Risks for the long run: A potential resolution of asset pricing

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility
and its use in estimating stochastic volatility models. *Journal of the Royal Statistical

R. J. Barro. Rare disasters and asset markets in the twentieth century. *The Quarterly

G. Bekaert and M. Hoerova. The VIX, the variance premium and stock market volatility.


This figure shows time series of TVRP, VRP and HRP from January 1996 to April 2016. TVRP is the total variance risk premium defined in Equation (5). VRP is the variance risk premium associated with the second moment of return, defined in Equation (13). HRP is the higher order risk premium associated with the third and higher moments of return, defined in Equation (16).
This figure shows sample autocorrelations of VRP and HRP from 1 lag to 24 lags. The gray lines represent the 95% confidence bands.
Figure 3: In-sample $R^2$ and Out-of-sample $R^2$ of Predictive Regressions

(a) In-sample $R^2$ of TVRP and Joint Regression

(b) Out-of-sample $R^2$ of TVRP and Joint Regression

The top panel plots the adjusted $R^2$ (in percentage) from return predictability regressions for the aggregate market portfolio based on VRP and HRP jointly (solid line) and TVRP (dashed line) as a function of forecasting horizon (in months). The bottom panel plots the corresponding out-of-sample $R^2$. 

46
This figure plots the model-implied covariance between market return in $r_{t+2h}$ and $q_t$ (left panel) and $\lambda_t$ (right panel), as a function of forecasting horizon $h$. In both panels, all other parameters are kept constant while we change the persistence parameter. On the left panel, we consider $\rho_q = 0.3$ (the dotted line), 0.5 (the solid line), and 0.7 (the dashed line). On the right panel, we consider $\rho_\lambda = 0.6$ (the dotted line), 0.9 (the solid line), and 0.95 (the dashed line). Horizons are measured in months.
This figure plots the model-implied predictive coefficients on VRP (left panel) and HRP (right panel), as a function of forecasting horizon $h$. In both panels, all other parameters are kept constant while we change the persistence parameter. On the left panel, we consider $\rho_q = 0.3$ (the dotted line), 0.5 (the solid line), and 0.7 (the dashed line). On the right panel, we consider $\rho_\lambda = 0.6$ (the dotted line), 0.9 (the solid line), and 0.95 (the dashed line). Horizons are measured in months.
This figure plots the model-implied $R^2$ of the predictive regression afforded by TVRP alone (the dotted line), and VRP and HRP jointly (the solid line), as a function of the predictive horizon $h$. Horizons are measured in months.
Table 1: Summary Statistics

Panel A: Summary Statistics of Option-implied Moments

<table>
<thead>
<tr>
<th></th>
<th>VIX$^2$</th>
<th>SVIX$^2$</th>
<th>KNS</th>
<th>IM3</th>
<th>IM4</th>
<th>IM2</th>
<th>IHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.72</td>
<td>4.38</td>
<td>-0.57</td>
<td>-0.69</td>
<td>0.25</td>
<td>4.93</td>
<td>-0.21</td>
</tr>
<tr>
<td>Std</td>
<td>4.28</td>
<td>3.79</td>
<td>0.95</td>
<td>1.25</td>
<td>0.62</td>
<td>4.62</td>
<td>0.37</td>
</tr>
<tr>
<td>Skewness</td>
<td>3.44</td>
<td>3.20</td>
<td>-6.11</td>
<td>-6.59</td>
<td>8.12</td>
<td>3.62</td>
<td>-6.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>19.31</td>
<td>17.03</td>
<td>53.93</td>
<td>61.60</td>
<td>87.79</td>
<td>21.09</td>
<td>58.23</td>
</tr>
<tr>
<td>Max</td>
<td>34.34</td>
<td>28.69</td>
<td>-0.02</td>
<td>-0.02</td>
<td>7.57</td>
<td>38.42</td>
<td>-0.01</td>
</tr>
<tr>
<td>Min</td>
<td>0.98</td>
<td>0.95</td>
<td>-10.37</td>
<td>-14.15</td>
<td>0.01</td>
<td>1.00</td>
<td>-4.09</td>
</tr>
</tbody>
</table>

Panel B: Summary Statistics of Realized Moments

<table>
<thead>
<tr>
<th></th>
<th>RTV</th>
<th>RSV</th>
<th>RKNS</th>
<th>RM3</th>
<th>RM4</th>
<th>RM2</th>
<th>RHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.81</td>
<td>3.80</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>3.81</td>
<td>0.00</td>
</tr>
<tr>
<td>Std</td>
<td>6.18</td>
<td>6.19</td>
<td>0.12</td>
<td>0.11</td>
<td>0.03</td>
<td>6.19</td>
<td>0.04</td>
</tr>
<tr>
<td>Skewness</td>
<td>5.92</td>
<td>5.98</td>
<td>-1.61</td>
<td>-3.38</td>
<td>11.57</td>
<td>5.89</td>
<td>-2.57</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>48.37</td>
<td>49.46</td>
<td>34.77</td>
<td>35.21</td>
<td>153.85</td>
<td>47.90</td>
<td>34.28</td>
</tr>
<tr>
<td>Max</td>
<td>63.06</td>
<td>63.67</td>
<td>0.80</td>
<td>0.56</td>
<td>0.47</td>
<td>62.83</td>
<td>0.23</td>
</tr>
<tr>
<td>Min</td>
<td>0.20</td>
<td>0.20</td>
<td>-0.94</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.20</td>
<td>-0.32</td>
</tr>
</tbody>
</table>

Panel C: Summary Statistics of Risk Premiums

<table>
<thead>
<tr>
<th></th>
<th>TVRP</th>
<th>VRP</th>
<th>M3RP</th>
<th>M4RP</th>
<th>HRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.91</td>
<td>1.12</td>
<td>-0.68</td>
<td>0.24</td>
<td>-0.21</td>
</tr>
<tr>
<td>Std</td>
<td>3.20</td>
<td>3.04</td>
<td>1.25</td>
<td>0.58</td>
<td>0.37</td>
</tr>
<tr>
<td>Skewness</td>
<td>-4.53</td>
<td>-3.46</td>
<td>-7.11</td>
<td>7.87</td>
<td>-7.00</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>38.26</td>
<td>28.52</td>
<td>71.04</td>
<td>83.10</td>
<td>69.23</td>
</tr>
<tr>
<td>Autocorr</td>
<td>0.42</td>
<td>0.38</td>
<td>0.58</td>
<td>0.59</td>
<td>0.58</td>
</tr>
<tr>
<td>Max</td>
<td>10.20</td>
<td>11.65</td>
<td>0.04</td>
<td>7.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Min</td>
<td>-28.72</td>
<td>-24.41</td>
<td>-14.72</td>
<td>0.00</td>
<td>-4.31</td>
</tr>
<tr>
<td>t-stats</td>
<td>4.45</td>
<td>5.76</td>
<td>-8.56</td>
<td>6.39</td>
<td>-8.83</td>
</tr>
</tbody>
</table>

Panel D: Correlation Matrix of the Risk Premiums

<table>
<thead>
<tr>
<th></th>
<th>TVRP</th>
<th>VRP</th>
<th>M3RP</th>
<th>M4RP</th>
<th>HRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVRP</td>
<td>1.00</td>
<td>0.99</td>
<td>0.50</td>
<td>-0.54</td>
<td>0.49</td>
</tr>
<tr>
<td>VRP</td>
<td>1.00</td>
<td>0.40</td>
<td>-0.45</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>M3RP</td>
<td>1.00</td>
<td>-0.99</td>
<td>1.00</td>
<td>-0.99</td>
<td></td>
</tr>
<tr>
<td>M4RP</td>
<td>1.00</td>
<td>-0.99</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRP</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports summary statistics of option-implied moments, realized moments, and risk premiums in Panel A, B, and C. The “t-stats” in Panel C are the t-statistics for testing the null hypothesis that the mean of the risk premium is equal to zero. Panel D reports the correlation matrix of the risk premiums. See Section 2 for more details on the definition of variables. All variables are denoted in percent per annum. The sample period extends from January 1996 to April 2016.
The table reports the estimated regression coefficient, $R^2$, adjusted $R^2$, and OOS $R^2$ of the predictability regressions from one to 24-month excess return on the S&P 500 index. For each horizon, we report predictive regressions results of different predictors: TVRP alone (1\textsuperscript{st} column), VRP alone (2\textsuperscript{nd} column), HRP alone (3\textsuperscript{rd} column), and VRP and HRP jointly (4\textsuperscript{th} column). The returns are observed monthly with the sample period ranging from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.
Table 3: Market Return Predictive Regressions: Sub-Sample without Realized Disaster

<table>
<thead>
<tr>
<th></th>
<th>1-month</th>
<th>3-month</th>
<th>6-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.92)</td>
<td>(1.48)</td>
</tr>
<tr>
<td>TVRP</td>
<td>0.35</td>
<td>0.54</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(2.07)</td>
<td>(3.18)</td>
<td></td>
</tr>
<tr>
<td>VRP</td>
<td>0.36</td>
<td>0.59</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>(2.18)</td>
<td>(3.53)</td>
<td>(3.34)</td>
</tr>
<tr>
<td>HRP</td>
<td>-3.87</td>
<td>-10.18</td>
<td>-11.10</td>
</tr>
<tr>
<td></td>
<td>(-3.71)</td>
<td>(-3.70)</td>
<td>(-2.26)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>3.05</td>
<td>2.78</td>
<td>4.09</td>
</tr>
<tr>
<td></td>
<td>(3.03)</td>
<td>(7.87)</td>
<td>(5.13)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>2.62</td>
<td>2.35</td>
<td>3.65</td>
</tr>
<tr>
<td></td>
<td>(2.97)</td>
<td>(3.03)</td>
<td>(4.22)</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>6.40</td>
<td>-4.05</td>
<td>-10.41</td>
</tr>
<tr>
<td></td>
<td>(7.10)</td>
<td>(5.66)</td>
<td>(4.26)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>9-month</th>
<th>12-month</th>
<th>24-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.48)</td>
<td>(0.42)</td>
</tr>
<tr>
<td>TVRP</td>
<td>0.88</td>
<td>0.87</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>(2.25)</td>
<td>(1.75)</td>
<td>(1.43)</td>
</tr>
<tr>
<td>VRP</td>
<td>0.89</td>
<td>0.88</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(1.87)</td>
<td>(1.45)</td>
</tr>
<tr>
<td>HRP</td>
<td>-8.80</td>
<td>-7.66</td>
<td>-3.88</td>
</tr>
<tr>
<td></td>
<td>(-1.56)</td>
<td>(-1.16)</td>
<td>(-0.41)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>2.58</td>
<td>1.74</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>(2.84)</td>
<td>(1.86)</td>
<td>(1.44)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>2.13</td>
<td>1.28</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(1.40)</td>
<td>(0.95)</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>-9.91</td>
<td>-12.23</td>
<td>-7.49</td>
</tr>
<tr>
<td></td>
<td>(-9.54)</td>
<td>(-11.82)</td>
<td>(-6.93)</td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, $R^2$, adjusted $R^2$, and OOS $R^2$ of the predictability regressions from one to 24-month excess return on the S&P 500 index. The sample period is from January 1996 to April 2016, excluding the crisis period: December 2007 to June 2009. For each horizon, we report predictive regressions results of different predictors: TVRP alone (1\textsuperscript{st} column), VRP alone (2\textsuperscript{nd} column), HRP alone (3\textsuperscript{rd} column), and VRP and HRP jointly (4\textsuperscript{th} column). Newey-West t-statistics are reported in the parenthesis.
Table 4: Out-of-sample CER Gains

<table>
<thead>
<tr>
<th>Panel A: with Short Selling Constraint</th>
<th>1-month</th>
<th>3-month</th>
<th>6-month</th>
<th>9-month</th>
<th>12-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma = 3 ) TVRP</td>
<td>2.90</td>
<td>1.71</td>
<td>0.89</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>VRP</td>
<td>3.51</td>
<td>2.03</td>
<td>1.22</td>
<td>0.31</td>
<td>-0.04</td>
</tr>
<tr>
<td>HRP+VRP</td>
<td>3.97</td>
<td>3.39</td>
<td>2.99</td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td>buy-and-hold</td>
<td>-1.13</td>
<td>-1.69</td>
<td>-1.73</td>
<td>-1.88</td>
<td>-1.90</td>
</tr>
<tr>
<td>( \gamma = 4 ) TVRP</td>
<td>2.17</td>
<td>1.28</td>
<td>0.67</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>VRP</td>
<td>2.63</td>
<td>1.53</td>
<td>0.92</td>
<td>0.23</td>
<td>-0.03</td>
</tr>
<tr>
<td>HRP+VRP</td>
<td>2.98</td>
<td>2.55</td>
<td>2.26</td>
<td>0.28</td>
<td>0.04</td>
</tr>
<tr>
<td>buy-and-hold</td>
<td>-1.25</td>
<td>-1.67</td>
<td>-1.71</td>
<td>-1.81</td>
<td>-1.80</td>
</tr>
<tr>
<td>( \gamma = 5 ) TVRP</td>
<td>1.74</td>
<td>1.03</td>
<td>0.54</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>VRP</td>
<td>2.11</td>
<td>1.23</td>
<td>0.74</td>
<td>0.19</td>
<td>-0.02</td>
</tr>
<tr>
<td>HRP+VRP</td>
<td>2.39</td>
<td>2.05</td>
<td>1.83</td>
<td>0.22</td>
<td>0.03</td>
</tr>
<tr>
<td>buy-and-hold</td>
<td>-1.36</td>
<td>-1.68</td>
<td>-1.72</td>
<td>-1.79</td>
<td>-1.76</td>
</tr>
</tbody>
</table>

Panel B: without Short Selling Constraint

| \( \gamma = 3 \) TVRP               | 3.61    | 2.75    | 2.18    | 0.02    | 0.92    |
| VRP                                  | 4.26    | 3.12    | 2.55    | -0.01   | 0.78    |
| HRP+VRP                              | 5.28    | 4.54    | 4.63    | 0.12    | -1.19   |
| buy-and-hold                         | -0.95   | -1.86   | -2.01   | -2.13   | -1.91   |
| \( \gamma = 4 \) TVRP               | 2.67    | 2.38    | 1.86    | 0.01    | 0.67    |
| VRP                                  | 3.13    | 2.60    | 2.09    | -0.03   | 0.56    |
| HRP+VRP                              | 3.77    | 3.59    | 3.56    | -0.14   | -1.10   |
| buy-and-hold                         | -1.12   | -1.79   | -1.92   | -1.99   | -1.80   |
| \( \gamma = 5 \) TVRP               | 2.07    | 1.92    | 1.57    | 0.01    | 0.51    |
| VRP                                  | 2.46    | 2.22    | 1.77    | -0.02   | 0.43    |
| HRP+VRP                              | 2.85    | 2.80    | 2.93    | -0.28   | -0.92   |
| buy-and-hold                         | -1.25   | -1.78   | -1.89   | -1.93   | -1.76   |

The table reports the annualized certainty equivalent return (CER) gain (in percent per annum) for a mean-variance investor who allocates between equities and risk-free bills using a predictive regression excess return forecast based on the predictor variable in the first column relative to the prevailing mean benchmark forecast. We report results for relative risk aversion coefficient \( \gamma \) of 3, 4 and 5. The equity weight is constrained to lie between 0 and 1.5 in Panel A and between -0.5 and 1.5 in Panel B. “Buy-and-hold” corresponds to the investor passively holding the market portfolio. We consider horizons of 1-month (third column), 3-month (fourth column), 9-month (fifth column), and 12-month (last column). We use non-overlapping returns so that the forecast horizon and rebalancing frequency coincide.
Table 5: Predictive Regressions: Returns of SMB, HML and WML at 3- and 9-Month Horizons

<table>
<thead>
<tr>
<th></th>
<th>SMB 3-month</th>
<th>HML 3-month</th>
<th>WML 3-month</th>
<th>SMB 9-month</th>
<th>HML 9-month</th>
<th>WML 9-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(-0.26)</td>
<td>(-0.29)</td>
<td>(-0.63)</td>
<td>(-0.93)</td>
<td>(-0.60)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>TVRP</td>
<td>-0.02</td>
<td>0.42</td>
<td>-0.05</td>
<td>0.05</td>
<td>0.39</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>(-0.20)</td>
<td>(1.68)</td>
<td>(-0.43)</td>
<td>(0.05)</td>
<td>(0.39)</td>
<td>(-0.93)</td>
</tr>
<tr>
<td>VRP</td>
<td>0.00</td>
<td>0.05</td>
<td>-0.15</td>
<td>0.19</td>
<td>-0.15</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(-0.04)</td>
<td>(0.33)</td>
<td>(-0.93)</td>
<td>(0.75)</td>
<td>(-0.93)</td>
<td>(-0.93)</td>
</tr>
<tr>
<td>HRP</td>
<td>-0.97</td>
<td>-1.13</td>
<td>4.86</td>
<td>4.25</td>
<td>6.43</td>
<td>8.20</td>
</tr>
<tr>
<td></td>
<td>(-1.79)</td>
<td>(-1.17)</td>
<td>(5.36)</td>
<td>(2.22)</td>
<td>(2.73)</td>
<td>(2.03)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.39</td>
<td>0.45</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>-0.41</td>
<td>-0.42</td>
<td>-0.03</td>
<td>-0.38</td>
<td>-0.39</td>
<td>-0.18</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>-4.67</td>
<td>-4.45</td>
<td>1.34</td>
<td>-4.48</td>
<td>-1.46</td>
<td>8.26</td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, $R^2$, adjusted $R^2$, and OOS $R^2$ for the predictive regressions for the excess 3-month and 9-month returns of the SMB, HML and WML factors. For each horizon, we report predictive regressions results of different predictors: TVRP alone (1st column), VRP alone (2nd column), HRP alone (3rd column), and VRP and HRP jointly (4th column). The returns are observed monthly with the sample period ranging from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.
Table 6: Predictive regressions with control variables at 3-month and 9-month horizons

<table>
<thead>
<tr>
<th></th>
<th>DY</th>
<th>EP</th>
<th>BM</th>
<th>TBL</th>
<th>DFY</th>
<th>LTY</th>
<th>NTIS</th>
<th>INFL</th>
<th>LTR</th>
<th>DFR</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: 3-month horizon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.36</td>
<td>0.06</td>
<td>-0.08</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(0.60)</td>
<td>(-3.46)</td>
<td>(-0.81)</td>
<td>(-0.55)</td>
<td>(0.64)</td>
<td>(-1.81)</td>
<td>(-1.51)</td>
<td>(-1.72)</td>
<td>(-0.80)</td>
<td>(-1.56)</td>
</tr>
<tr>
<td>VRP</td>
<td>1.02</td>
<td>1.08</td>
<td>1.04</td>
<td>1.01</td>
<td>1.00</td>
<td>1.02</td>
<td>0.89</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>(4.26)</td>
<td>(4.29)</td>
<td>(4.10)</td>
<td>(3.80)</td>
<td>(3.54)</td>
<td>(3.89)</td>
<td>(3.84)</td>
<td>(3.93)</td>
<td>(3.74)</td>
<td>(3.51)</td>
<td>(4.11)</td>
</tr>
<tr>
<td>HRP</td>
<td>-3.48</td>
<td>-6.01</td>
<td>-4.39</td>
<td>-4.53</td>
<td>-5.55</td>
<td>-4.67</td>
<td>-5.26</td>
<td>-4.97</td>
<td>-4.80</td>
<td>-4.72</td>
<td>-4.60</td>
</tr>
<tr>
<td></td>
<td>(-2.55)</td>
<td>(-5.65)</td>
<td>(-4.25)</td>
<td>(-3.98)</td>
<td>(-4.28)</td>
<td>(-4.34)</td>
<td>(-3.76)</td>
<td>(-4.11)</td>
<td>(-3.94)</td>
<td>(-3.77)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.09</td>
<td>0.03</td>
<td>0.24</td>
<td>-0.36</td>
<td>-1.05</td>
<td>-0.83</td>
<td>0.64</td>
<td>0.41</td>
<td>0.07</td>
<td>0.18</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(0.82)</td>
<td>(3.04)</td>
<td>(-1.08)</td>
<td>(-0.57)</td>
<td>(-1.31)</td>
<td>(1.22)</td>
<td>(0.31)</td>
<td>(0.56)</td>
<td>(1.23)</td>
<td>(-2.41)</td>
</tr>
<tr>
<td>Panel B: 9-month horizon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.11</td>
<td>0.20</td>
<td>-0.25</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.07</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(2.67)</td>
<td>(0.81)</td>
<td>(-3.20)</td>
<td>(0.05)</td>
<td>(-0.69)</td>
<td>(0.78)</td>
<td>(-1.24)</td>
<td>(-0.79)</td>
<td>(-0.97)</td>
<td>(-0.20)</td>
<td>(-0.76)</td>
</tr>
<tr>
<td>VRP</td>
<td>1.15</td>
<td>1.31</td>
<td>1.22</td>
<td>1.10</td>
<td>1.12</td>
<td>1.14</td>
<td>0.64</td>
<td>1.19</td>
<td>1.13</td>
<td>1.09</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>(3.57)</td>
<td>(4.33)</td>
<td>(3.82)</td>
<td>(3.29)</td>
<td>(3.34)</td>
<td>(3.54)</td>
<td>(1.77)</td>
<td>(2.84)</td>
<td>(3.22)</td>
<td>(2.92)</td>
<td>(2.86)</td>
</tr>
<tr>
<td></td>
<td>(-2.06)</td>
<td>(-6.38)</td>
<td>(-3.87)</td>
<td>(-3.51)</td>
<td>(-3.08)</td>
<td>(-3.63)</td>
<td>(-3.04)</td>
<td>(-3.53)</td>
<td>(-3.52)</td>
<td>(-3.25)</td>
<td>(-3.41)</td>
</tr>
<tr>
<td>Other</td>
<td>0.28</td>
<td>0.08</td>
<td>0.82</td>
<td>-1.41</td>
<td>0.37</td>
<td>-2.20</td>
<td>2.70</td>
<td>-4.04</td>
<td>0.13</td>
<td>0.64</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(2.87)</td>
<td>(1.00)</td>
<td>(3.48)</td>
<td>(-1.35)</td>
<td>(0.08)</td>
<td>(-1.10)</td>
<td>(1.67)</td>
<td>(-0.98)</td>
<td>(0.55)</td>
<td>(2.01)</td>
<td>(-2.30)</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>6.09</td>
<td>-7.57</td>
<td>25.59</td>
<td>11.91</td>
<td>-4.25</td>
<td>11.45</td>
<td>-5.79</td>
<td>7.07</td>
<td>5.42</td>
<td>7.30</td>
<td>29.92</td>
</tr>
</tbody>
</table>

This table reports regression results of HRP, VRP and control variables at the 3-month and 9-month horizons in Panel A and B. In each column of the table, we add one control variable to the regression of HRP and VRP. The “Other” variable is specified in the first row of the table. The control variables are defined in Section 5.1. The first ten variables are obtained from Amit Goyal’s website. The last variable is obtained from David Rapach’s website. The Newy-West t-statistics are reported in the parenthesis. We report $R^2$, adjusted $R^2$, and out-of-sample $R^2$ in the last three rows. The sample period is from January 1996 to April 2016.
Table 7: Market Return Predictive Regressions Using Predicted Value of Realized Variances

<table>
<thead>
<tr>
<th></th>
<th>1-month</th>
<th>3-month</th>
<th>6-month</th>
<th>9-month</th>
<th>12-month</th>
<th>24-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(-0.18)</td>
<td>(-0.25)</td>
<td>(0.13)</td>
<td>(-0.23)</td>
<td>(-0.15)</td>
<td>(-0.23)</td>
</tr>
<tr>
<td>TVRP&lt;sub&gt;f&lt;/sub&gt;</td>
<td>0.11</td>
<td>0.30</td>
<td>0.29</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(2.78)</td>
<td>(3.55)</td>
<td>(1.65)</td>
<td>(1.64)</td>
<td>(1.64)</td>
</tr>
<tr>
<td>VRP&lt;sub&gt;f&lt;/sub&gt;</td>
<td>0.11</td>
<td>0.30</td>
<td>0.29</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(2.78)</td>
<td>(3.55)</td>
<td>(1.65)</td>
<td>(1.64)</td>
<td>(1.64)</td>
</tr>
<tr>
<td>HRP&lt;sub&gt;f&lt;/sub&gt;</td>
<td>-0.19</td>
<td>-0.23</td>
<td>-0.18</td>
<td>-8.13</td>
<td>-8.13</td>
<td>-10.11</td>
</tr>
<tr>
<td></td>
<td>(-0.59)</td>
<td>(-0.42)</td>
<td>(-0.42)</td>
<td>(-3.30)</td>
<td>(-3.30)</td>
<td>(-3.30)</td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1.16</td>
<td>2.83</td>
<td>0.19</td>
<td>0.36</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>Adj. R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.75</td>
<td>2.42</td>
<td>0.34</td>
<td>-0.24</td>
<td>-0.06</td>
<td>-0.06</td>
</tr>
<tr>
<td>OOS R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-9.14</td>
<td>-6.21</td>
<td>-20.03</td>
<td>-7.19</td>
<td>-5.41</td>
<td>-25.39</td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, R<sup>2</sup>, adjusted R<sup>2</sup>, and OOS R<sup>2</sup> of the predictability regressions from one to 24-month excess returns on the S&P 500 index. For each horizon, we report predictive regressions results of different predictors: TVRP<sub>f</sub> alone, VRP<sub>f</sub> alone, HRP<sub>f</sub> alone, and VRP<sub>f</sub> and HRP<sub>f</sub> jointly. TVRP<sub>f</sub> is calculated as the difference between VIX<sup>2</sup> and the predicted value of RTV. VRP<sub>f</sub> is the difference between IM2 and the predicted value of RM2. HRP<sub>f</sub> is the difference between TVRP<sub>f</sub> and VRP<sub>f</sub>. We use a moving window of 60 months to get one-step-ahead forecast of RTV and RM2 using their lagged values and RM3. See more details in Section 5.2. The sample period extends from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.
Table 8: Summary Statistics of Risk Premia Using High Frequency Realized Moments

Panel A: Summary Statistics of the Intraday Realized Moments and the Corresponding Risk Premia

<table>
<thead>
<tr>
<th></th>
<th>RM2^{5-min}</th>
<th>RM3^{5-min}</th>
<th>RM4^{5-min}</th>
<th>TVRP^{5-min}</th>
<th>VRP^{5-min}</th>
<th>M3RP^{5-min}</th>
<th>M4RP^{5-min}</th>
<th>HRP^{5-min}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.93</td>
<td>-1.5 \times 10^{-3}</td>
<td>2.83 \times 10^{-4}</td>
<td>1.78</td>
<td>2.00</td>
<td>-0.69</td>
<td>0.25</td>
<td>-0.21</td>
</tr>
<tr>
<td>Std</td>
<td>4.86</td>
<td>0.01</td>
<td>0.00</td>
<td>2.64</td>
<td>2.60</td>
<td>1.24</td>
<td>0.61</td>
<td>0.36</td>
</tr>
<tr>
<td>Skewness</td>
<td>7.70</td>
<td>-7.15</td>
<td>11.79</td>
<td>-3.84</td>
<td>-1.82</td>
<td>-6.56</td>
<td>8.11</td>
<td>-6.35</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>83.57</td>
<td>65.68</td>
<td>157.68</td>
<td>54.24</td>
<td>36.01</td>
<td>60.96</td>
<td>87.51</td>
<td>57.55</td>
</tr>
<tr>
<td>Max</td>
<td>60.26</td>
<td>0.03</td>
<td>0.02</td>
<td>16.06</td>
<td>17.69</td>
<td>-0.02</td>
<td>7.54</td>
<td>-0.01</td>
</tr>
<tr>
<td>Min</td>
<td>0.34</td>
<td>-0.14</td>
<td>0.00</td>
<td>-25.88</td>
<td>-21.84</td>
<td>-14.02</td>
<td>0.01</td>
<td>-4.05</td>
</tr>
</tbody>
</table>

Panel B: Correlation Matrix of the Daily and Intraday Realized Moments

<table>
<thead>
<tr>
<th></th>
<th>RM2^{5-min}</th>
<th>RM3^{5-min}</th>
<th>RM4^{5-min}</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM2</td>
<td>0.96</td>
<td>-0.37</td>
<td>0.71</td>
</tr>
<tr>
<td>RM3</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>RM4</td>
<td>0.92</td>
<td>-0.51</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Panel A reports summary statistics of the intraday realized moments and their corresponding risk premiums. Panel B reports the correlation matrix of the daily and intraday realized higher moments. The intraday realized moments are calculated from 5-min returns of the S&P 500 cash index. All variables are annualized and denoted in percent. The sample period extends from January 1996 to April 2016.
<table>
<thead>
<tr>
<th></th>
<th>1-month</th>
<th>3-month</th>
<th>6-month</th>
<th>9-month</th>
<th>12-month</th>
<th>24-month</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>-0.01 (-2.09)</td>
<td>-0.01 (-1.44)</td>
<td>-0.01 (-0.87)</td>
<td>-0.01 (-0.46)</td>
<td>-0.02 (-0.40)</td>
<td>-0.03 (-0.46)</td>
</tr>
<tr>
<td><strong>TVRP(^{5\text{-min}})</strong></td>
<td>0.38 (4.27)</td>
<td>0.80 (6.41)</td>
<td>0.97 (4.18)</td>
<td>5.12 (2.88)</td>
<td>4.73 (2.09)</td>
<td>8.32 (2.88)</td>
</tr>
<tr>
<td><strong>VRP(^{5\text{-min}})</strong></td>
<td>0.39 (3.89)</td>
<td>0.39 (3.70)</td>
<td>0.85 (7.11)</td>
<td>7.89 (2.88)</td>
<td>7.51 (2.88)</td>
<td>3.14 (2.88)</td>
</tr>
<tr>
<td><strong>HRP(^{5\text{-min}})</strong></td>
<td>-0.14 (-0.15)</td>
<td>-0.24 (-0.58)</td>
<td>-1.50 (-0.58)</td>
<td>-15.04 (-3.23)</td>
<td>-10.04 (-3.69)</td>
<td>-12.51 (-3.69)</td>
</tr>
<tr>
<td><strong>R(^2)</strong></td>
<td>5.12 5.35 0.01</td>
<td>5.39 5.87 0.53</td>
<td>9.43 5.38 3.74</td>
<td>11.03 6.89 3.74</td>
<td>10.27 6.50 3.33</td>
<td>23.55 7.43 3.74</td>
</tr>
<tr>
<td><strong>Adj. R(^2)</strong></td>
<td>4.73 4.96 -0.40</td>
<td>4.60 7.51 8.34</td>
<td>8.67 4.98 6.50</td>
<td>10.27 6.50 3.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OOS R(^2)</strong></td>
<td>8.32 8.36 -7.14</td>
<td>2.06 3.14 6.18</td>
<td>-25.53 3.14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, $R^2$, adjusted $R^2$, and OOS $R^2$ of the predictability regressions from one to 24-month excess return on the S&P 500 index. For each horizon, we report predictive regressions results of different predictors: TVRP\(^{5\text{-min}}\), VRP\(^{5\text{-min}}\), HRP\(^{5\text{-min}}\), and VRP\(^{5\text{-min}}\) and HRP\(^{5\text{-min}}\) jointly. VRP\(^{5\text{-min}}\) and HRP\(^{5\text{-min}}\) are calculated using 5-min intraday return of the S&P 500 cash index. The sample extends from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.
Table 10: Calibrated Parameters

Panel A: Baseline Parameters

<table>
<thead>
<tr>
<th>Preferences</th>
<th>$\delta$</th>
<th>$\gamma$</th>
<th>$\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta c_{t+1}$</td>
<td>$\mathbb{E}[\Delta c_{t+1}]$</td>
<td>$\rho_c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$8 \times 10^{-4}$</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{t+1}$</td>
<td>$\mathbb{E}[\sigma^2_{t+1}]$</td>
<td>$\rho_{\sigma}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$3 \times 10^{-5}$</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>$q_{t+1}$</td>
<td>$\mathbb{E}[q]$</td>
<td>$\rho_q$</td>
<td>$\phi_q$</td>
</tr>
<tr>
<td></td>
<td>$5 \times 10^{-7}$</td>
<td>0.6</td>
<td>$3.6 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\lambda_{t+1}$</td>
<td>$\mathbb{E}[\lambda^2_{t+1}]$</td>
<td>$\rho_{\lambda}$</td>
<td>$\phi_{\lambda}$</td>
</tr>
<tr>
<td></td>
<td>0.05/24</td>
<td>0.95</td>
<td>0.025</td>
</tr>
<tr>
<td>$J$</td>
<td>$\mathbb{E}_v[J]$</td>
<td>$\mu_j$</td>
<td>$\nu_j$</td>
</tr>
<tr>
<td></td>
<td>-2%</td>
<td>0.02</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Panel B: VRP and HRP’s Loadings on $q_t$ and $\lambda_t$

<table>
<thead>
<tr>
<th>VRP</th>
<th>$c_q^{\text{vrp}}$</th>
<th>$c_{\lambda}^{\text{vrp}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$3.1 \times 10^5$</td>
<td>5.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HRP</th>
<th>$c_q^{\text{hrp}}$</th>
<th>$c_{\lambda}^{\text{hrp}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>-0.52</td>
</tr>
</tbody>
</table>

Panel A of this table reports the baseline parameter values used to produce Figure 4, Figure 5, and Figure 6. These baseline parameters are converted from typical values used in Bollerslev, Tauchen, and Zhou (2009), Drechsler and Yaron (2010), Guo, Sha, Wang, and Zhou (2018), and Wachter (2013) to half-a-month quantities. Exceptions are $\rho_{\sigma}$, $\rho_q$, and $\phi_q$. Panel B reports the loadings on $q_t$ and $\lambda_t$ in VRP and HRP. The loadings are given by Equation (39) and (40), calculated with the baseline parameters in Panel A of this table. See Appendix C.6 for details on the parameter calibration.
Appendices

A Proofs

Proof of Proposition 1. Using Equation (3), (6), (7) and (8), it holds that

\[ \text{IM}_4 = \frac{1}{T-t} E^Q \left[ g_4(r(t, T)) \right] = \frac{1}{T-t} E^Q \left[ -6 \left( g_v(r(t, T)) - g_{sv}(r(t, T)) \right) - 4g_{kns}(r(t, T)) \right] \\
= - \frac{6}{T-t} E^Q [g_v(r(t, T)) - g_{sv}(r(t, T))] - \frac{4}{T-t} E^Q [g_{kns}(r(t, T))] \\
= -6(VIX_t^2 - 4SVIX_t^2) - \text{KNS}_t. \]

Similarly,

\[ \text{IM}_3 = \frac{1}{T-t} E^Q \left[ g_3(r(t, T)) \right] = \frac{1}{T-t} E^Q \left[ g_{kns}(r) - \frac{1}{2} g_4(r) \right] \\
= \frac{1}{T-t} E^Q [g_{kns}(r)] - \frac{1}{2(T-t)} E^Q [g_4(r(t, T))] \\
= \text{KNS}_t - \frac{1}{2} \text{IM}_4. \]

\[ \text{IM}_2 = \frac{1}{T-t} E^Q \left[ g_2(r(t, T)) \right] = \frac{1}{T-t} E^Q \left[ g_v(r(t, T)) - \frac{1}{3} g_3(r(t, T)) - \frac{1}{12} g_4(r(t, T)) \right] \\
= \text{VIX}_t^2 - \frac{1}{3} \text{IM}_3 - \frac{1}{12} \text{IM}_4. \]

\qed
Proof of Proposition 2. By the definition of TVRP of Equation (5), we have

\[
\text{TVRP}_t := \text{VIX}_t^2 - \text{RTV}_t = \frac{1}{T-t} \mathbb{E}_t^Q \left[ g_v(r(t, T)) \right] - \frac{1}{T-t} \sum_{i=1}^{n} 2(e^{r_i} - 1 - r_i)
\]

\[
= \frac{1}{T-t} \left\{ \mathbb{E}_t^Q [r(t, T)^2 + \frac{1}{3} r(t, T)^3 + \frac{1}{12} r(t, T)^4 + o(r(t, T)^4)]
\right. \\
- \left. \sum_{i=1}^{n} \left( r_i^2 + \frac{1}{3} r_i^3 + \frac{1}{12} r_i^4 + o(r_i^4) \right) \right\}
\]

\[
= \frac{1}{T-t} \left\{ \text{VIX}_t^2 + \frac{1}{3} \text{IM}_3_t + \frac{1}{12} \text{IM}_4_t - \text{RM}_2_t - \frac{1}{3} \text{RM}_3_t - \frac{1}{12} \text{RM}_4_t \right\}
\]

\[
= \text{VRP}_t + \text{HRP}_t.
\]

\[\square\]

B Additional Robustness Checks

B.1 Bakshi, Kapadia, and Madan (2003) Contracts

In this section, we use an alternative method to decompose TVRP with the cubic and quartic contracts defined in Bakshi, Kapadia, and Madan (2003). Bakshi, Kapadia, and Madan (2003) show that the fair values of the cubic and quartic contracts, \( W_t := \mathbb{E}_t^Q [r(t, T)^3] \) and \( X_t := \mathbb{E}_t^Q [r(t, T)^4] \), are given by,

\[
W_t = \int_{F_t}^{\infty} \frac{6 \log(K/F_t) - 3(\log(K/F_t))^2}{K^2} C_t(K, T) dK
\]

\[
- \int_{0}^{F_t} \frac{6 \log(F_t/K) + 3(\log(F_t/K))^2}{K^2} P_t(K, T) dK,
\]

\[
X_t = \int_{F_t}^{\infty} \frac{12(\log(F_t/K))^2 - 4(\log(K/F_t))^3}{K^2} C_t(K, T) dK
\]

\[
+ \int_{0}^{F_t} \frac{12(\log(F_t/K))^2 + 4(\log(F_t/K))^3}{K^2} P_t(K, T) dK.
\]

Similar to Proposition 2, we can replicate VRP and HRP using the cubic and quartic contracts in Bakshi, Kapadia, and Madan (2003) and their corresponding realized measures:
TVRP = VRP_{BKM} + HRP_{BKM}. VRP_{BKM} and HRP_{BKM} are defined as,

\[
HRP_{BKM} = \frac{1}{3} \left( W_t - RM_{3t}^{BKM} \right) + \frac{1}{12} \left( X_t - RM_{4t}^{BKM} \right),
\]

\[
VRP_{BKM} = TVRP - HRP_{BKM},
\]

where TVRP is given by Equation (5), RM_{3t}^{BKM} = \sum_{i=1}^{n} r_{3t}^{i} and RM_{4t}^{BKM} = \sum_{i=1}^{n} r_{4t}^{i}. To calculate W_t and X_t, we use the same option price data at the end of each month as in Section 3. The realized moments are calculated from daily log returns in each month.

Table B1 shows that the predictive regression results using TVRP, VRP_{BKM}, and HRP_{BKM} are very similar to those in Table 2, in terms of t-statistics, in-sample R^2, OOS R^2, and the improvement of predictive power in multivariate over univariate regressions.

### B.2 Schneider and Trojani (2015) Contracts

In this section, we implement another decomposition of TVRP using the moment risk premia calculated from the contracts introduced by Schneider and Trojani (2015) and Schneider and Trojani (2017).

Let \( \Phi : \mathbb{R}_{+} \to \mathbb{R} \) be a scalar function, assumed to be twice differentiable everywhere. Define a payoff function \( h(F_T, F_0) \) as

\[
h(F_T, F_0) := \Phi(F_T) - \Phi(F_0) - \Phi'(F_0)(F_T - F_0).
\]

One can show that the payoff function \( h \) has the Aggregation Property II as defined by Bondarenko (2014). The forward fair value of the payoff function \( h \) is given by

\[
\mathbb{E}^Q[h] = \int_0^\infty \Phi''(K)\Theta(K) dK.
\]

Schneider and Trojani (2015) consider a particular \( \Phi \) function

\[
\Phi_q(x) := \frac{x^q - 1}{q(q - 1)}.
\]
Based on this particular $\Phi_q$ function, the payoff $h_q$ is given by

$$h_q(F_T, F_0) := \Phi_q(F_T) - \Phi_q(F_0) - \Phi_q'(F_0)(F_T - F_0).$$

Let $r$ denote the log futures return, $r = \log(F_T) - \log(F_0)$. The Schneider and Trojani (2015) variance swap contract has floating leg

$$H_2(r, q) := \frac{h_q}{F_0^q} = \frac{e^{rq} - 1 - q(e^r - 1)}{q(q - 1)} = \frac{1}{2} r^2 + \sum_{k=3}^{\infty} A(k, q) r^k,$$

where

$$A(k, q) = \frac{q^{k-1} - 1}{k!(q - 1)}.$$

Schneider and Trojani (2015) show that the fixed leg of this variance swap is given by

$$\mathbb{E}^Q[H_2] = \frac{1}{F_0^2} \int_0^{\infty} (K/F_0)^{q-2} \Theta(K) dK.$$

The Schneider and Trojani (2015) variance swap is a generalized variance swap, since it has a leading exposure to the variance of returns. In the special case that $q = 2$, the swap is a simple variance swap introduced by Martin (2017). In this case $\frac{1}{T-t} \mathbb{E}^Q[H_2(r(t, T), 2)]$ coincides with the squared SVIX index.

To further calculate the higher moments, Schneider and Trojani (2017) introduce a series of simple swaps. Similar to the idea of computing moments from a moment generating function, the floating leg of a simple swap of order $j$ is the $(j - 2)$-th derivative of $H_2$ with respect to $q$,

$$H_j(r, q) := \frac{\partial^{j-2}H_2}{\partial q^{j-2}} = \sum_{k=j}^{\infty} A_j(k, q) r^k, \quad j = 2, 3, \ldots,$$
where $A_j(k,q) = \frac{\partial^{j-2} A(k,q)}{\partial q^{j-2}}$. The fixed leg of the simple swap is given by

$$\text{IH}_j(q) := \mathbb{E}^Q[H_j] = \frac{1}{F_0^j} \int_0^\infty \left( \log \left( \frac{K}{F_0} \right) \right)^{j-2} \left( \frac{K}{F_0} \right)^{q-2} \Theta(K)dK, \quad j = 2, 3, \ldots$$

Observe that the floating leg $H_j$ of a $j$-th order simple swap has a leading contribution to moments of order $j$ and a zero contribution to moments of order less than $j$.

One important feature of these contracts is that it can be dynamically replicated in a way that is robust to the sampling frequency of returns, i.e.,

$$\mathbb{E}^Q[H_j(r(t,T))] = \mathbb{E}^Q \left[ \sum_{i=1}^n H_j(r_i) \right], \quad \forall j \geq 2,$$

for any partition $t_1, t_2, \ldots, t_n$ of $[t, T]$. This is known as the Aggregation Property proposed in Kozhan, Neuberger, and Schneider (2013) and Bondarenko (2014).8

The options-implied second, third and fourth moments of the Schneider and Trojani (2015) contract are:

$$\text{IH}_2 = \mathbb{E}^Q[H_2] = \int_0^\infty \frac{1}{\sqrt{F_0 K^3}} \Theta(K)dK,$$

$$\text{IH}_3 = \mathbb{E}^Q[H_3] = \int_0^\infty \frac{1}{\sqrt{F_0 K^3}} \log \left( \frac{K}{F_0} \right) \Theta(K)dK,$$

$$\text{IH}_4 = \mathbb{E}^Q[H_4] = \int_0^\infty \frac{1}{\sqrt{F_0 K^3}} \left( \log \left( \frac{K}{F_0} \right) \right)^2 \Theta(K)dK.$$

8The Schneider and Trojani (2015) contracts are not the only type of contract that has this property. For example, the realized skewness in Neuberger (2012) also has the aggregation property. We choose the Schneider and Trojani (2015) contracts in this section in particular because they do not require dynamic option trading in the floating leg of the skew swaps. These qualifications do not affect our return predictability exercises.
The realized second, third and fourth moments are:

\[
RH_2 = \sum_{i=1}^{n} -4(e_{r,i}^{\alpha/2} - 1) + 2(e_{r,i}^{\beta} - 1) = \frac{1}{2}r_i^2 + \frac{1}{4}r_i^3 + \frac{7}{96}r_i^4 + o(r_i^4),
\]

\[
RH_3 = \sum_{i=1}^{n} 4(e_{r,i}^{\alpha/2} + (e_{r,i}^{\beta} - 1)) = \frac{1}{6}r_i^3 + \frac{1}{12}r_i^4 + o(r_i^4),
\]

\[
RH_4 = \sum_{i=1}^{n} 4r_i^2e_{r,i}^{\alpha/2} + 32(e_{r,i}^{\alpha/2} - 1) - 16(e_{r,i}^{\beta} - 1) = \frac{1}{12}r_i^4 + o(r_i^4).
\]

We can decompose TVRP using the Schneider and Trojani (2017) moment risk premiums

\[
TVRP = VRP^H + HRP^H,
\]

where

\[
MRP_4^H := 12\left(\text{IH}_4 - RH_4\right),
\]

\[
MRP_3^H := 6\left(\text{IH}_3 - RH_3\right) - \frac{1}{2}MRP_4^H,
\]

\[
HRP^H := \frac{1}{2}MRP_3^H + \frac{7}{48}MRP_4^H,
\]

\[
VRP^H := TVRP - HRP^H.
\]

The options-implied Schneider and Trojani (2017) moments are calculated from the same option price data at the end of each month, and the realized counterparts are calculated from the same daily return data within each month, as in Section 3.

As a matter of fact, these Schneider and Trojani (2015) contracts allow us to calculate the higher order risk premium and variance risk premium truncated at any arbitrary order \(k\). Recall that TVRP = VRP + \(\sum_{p=3}^{\infty} a_p MRP_p\), in which the \(p\)-th order risk premium is given by \(MRP_p(q) = A_p(k,q)^{-1} (\text{IH}_p(q) - \sum_{i=1}^{n} H_p(r_i,q))\). To truncate at order \(k\), one simply needs to start from \(H_k\) and \(\text{IH}_k\), and then go backwards.

Table B2 reports the predictive regression results for S&P 500 index return using TVRP, \(VRP^H\), and \(HRP^H\).\(^9\) The prediction results are qualitatively similar to Table ____________

\(^9\)Schneider and Trojani (2017) give an extensive discussion on how to choose \(q\) optimally. They
2. There is significant improvement in in-sample and OOS $R^2$'s when TVRP is decomposed into the VRP$^H$ and HRP$^H$, despite that some of the $R^2$'s and $t$-statistics in these predictive regressions are slightly smaller than those in Table 2.

B.3 Jump- or Skewness-Related Predictors

The higher order risk premium, dominated by the third moment risk premium, closely mimics jump and asymmetric return measures. To show that VRP and HRP jointly give better forecasting performance, we run bivariate regressions of VRP joint with another jump- or skewness-related predictor proposed in the literature. In particular, we consider four candidate predictors: (1) implied jump measure, JTIX, as in Du and Kapadia (2012), (2) realized signed jump, RSJ, as in Guo, Sha, Wang, and Zhou (2018), (3) implied skewness, ISK, proposed in Bakshi, Kapadia, and Madan (2003), and (4) realized skewness calculated based on daily returns, RSK.

The implied jump measure (JTIX) measure proposed by Du and Kapadia (2012) is constructed as

$$JTIX_t = V_t^{BKM} - VIX_t^2,$$

$$V_t^{BKM} = \int_{F_t}^{\infty} \frac{2(1 + \log(F_t/K))}{K^2} C_t(K,T) dK + \int_0^{F_t} \frac{2(1 + \log(F_t/K))}{K^2} P_t(K,T) dK.$$

The realized signed jump measure (RSJ) proposed by Guo, Sha, Wang, and Zhou (2018) is calculated as,

$$RSJ_t = \sum_{i=1}^{n} 1_{\{r_i > 0\}} RJV_i - 1_{\{r_i < 0\}} RJV_i,$$

where RJV$_i$ is the realized jump volatility, that is, the realized volatility due to jumps of day $i$ if a jump is detected at day $i$.

The implied skewness (ISK) proposed in Bakshi, Kapadia, and Madan (2003) has propose to use $q = 1/2$ to mitigate higher moments contamination. In our case, since we isolate each moment using linear combinations of the Schneider and Trojani (2015) contracts, the particular choice of $q$ should not make a qualitative difference.
been used by Stilger, Kostakis, and Poon (2016), Conrad, Dittmar, and Ghysels (2013), and Chang, Christoffersen, and Jacobs (2013) to predict cross-section variation of stock returns. ISK is constructed as,

\[
ISK_t := \frac{\mathbb{E}^Q[r(t, T)^3]}{\left(\mathbb{E}^Q[r(t, T)^2]\right)^{3/2}} = \frac{1}{(V^{\text{BK0}})^{3/2}} \left\{ \int_{F_t}^{\infty} \frac{6 \log(K/F_t) - 3(\log(K/F_t))^2}{K^2} C_t(T, K) dK 
- \int_0^{F_t} \frac{6 \log(K/F_t) + 3(\log(K/F_t))^2}{K^2} P_t(T, K) dK \right\}.
\]

(35)

Amaya, Christoffersen, Jacobs, and Vasquez (2015) find a negative relation between the realized skewness and next week’s stock returns in a cross-section analysis. The realized skewness (RSK) is constructed as

\[
RSK_t = \frac{\sum_{i=1}^{n} r_i^3}{\left(\sum_{i=1}^{n} r_i^3\right)^{3/2}}.
\]

(36)

Table B3 reports the prediction results using VRP and the aforementioned predictors. In each panel, we replace HRP in Equation (18) with one alternative predictor over different horizons. Results in Table 2 show that the coefficients on VRP and HRP in the joint regressions are significant for all horizons, and the OOS $R^2$’s are positive for all horizons. However, none of the alternative HRP replacement considered in Table B3 has all significant coefficient as well as all positive OOS $R^2$ for every horizon. The performance of JTIX is comparable to HRP, but the OOS $R^2$ is negative at the very short 1-month horizon. RSJ produces decent predictability up to a quarter, but fails to forecast in the longer horizons. Both ISK and RSK contain little predictive power for market returns. None of the coefficients on ISK or RSK is statistically significant at any horizon.

While it is difficult to ascertain the dominance of HRP over all potential alternative predictors at all horizons, evidence here seems to suggest that our measure of HRP has stronger predictive power than the ad hoc jump or tail measures in the prevailing literature, both in-sample and out-of-sample, for both short and longer horizons.
C Model Solution

In this section, we provide more details on the derivation of model-implied VRP and HRP and how we calibrate the parameters.

C.1 Solving the Model

We first solve $A_0$, $A_1$, $\kappa_0$, $\kappa_1$ of the model. Plugging the pricing kernel given by Equation (21) and market return (23) into the Euler equation yields

$$\mathbb{E}_t \left[ \exp \left( \theta \log \delta - \theta \left( \frac{1}{\psi} - 1 \right) \Delta c_{t+1} + \theta \kappa_0 + \theta \kappa_1 w_{t+1} - \theta w_t \right) \right] = 1.$$

Substitute the expression for $w_t$,

$$\exp \left( \theta \log \delta + \theta (\kappa_1 - 1) A_0 + \theta \kappa_0 - \theta A'_1 Y_t \right) \mathbb{E}_t \left[ \exp \left( (\theta(1 - \frac{1}{\psi})c + \theta \kappa_1 A_1)'Y_{t+1} \right) \right] = 1.$$

Define $u \equiv \theta(1 - \frac{1}{\psi})c + \theta \kappa_1 A_1$. The conditional expectation in Equation (37) has an exponential affine form

$$\mathbb{E}_t[\exp(u'Y_{t+1})] = \exp(\alpha + \beta'Y_t),$$

where

$$\left\{ \begin{array}{l}
\alpha = \mu' u + \frac{1}{2} u'H_0 u, \\
\beta = F' u + \frac{1}{2} u'H_1 u + e_\lambda (\mathbb{E}_u[\exp(u') - 1]).
\end{array} \right.$$  

Take logarithm on both sides of Equation (37) and get,

$$\theta \log \delta + \theta (\kappa_1 - 1) A_0 + \theta \kappa_0 + \alpha + (\beta - \theta A_1)'Y_t = 0.$$
The constant and coefficient on $Y_t$ must be zero, implying

$$
\begin{align*}
A_0 &= \frac{\theta \log \delta + \theta \kappa_0 + \alpha}{\theta (1 - \kappa_1)} \\
A_1 &= \frac{1}{\theta} \beta.
\end{align*}
$$

Having solved $A_0, A_1$ as functions of $\kappa_0, \kappa_1$, we can proceed to solve $\kappa_0, \kappa_1$. According to Campbell and Shiller (1988),

$$
\begin{align*}
\kappa_0 &= \log(1 + \exp(E[w_t])) - \kappa_1 E[w_t], \\
\kappa_1 &= \frac{\exp(E[w_t])}{1 + \exp(E[w_t])},
\end{align*}
$$

where $E[w_t]$ is the unconditional expectation of $w_t$ given by

$$
E[w_t] = A_0 + A_1' E[Y_t].
$$

Since $Y_t$ is a stationary process, $E[Y_t]$ can be solved by $(I - F)E[Y_t] = \mu + e_c E[v](e' \Lambda E[Y_t])$.

**C.2 Risk Free Rate**

We next derive the model-implied risk free rate. Substitute $r_{t+1}$ for $r_{f,t}$ in the Euler Equation (6.1) and get $r_{f,t} = -\log E_t[\exp(m_{t+1})]$. The expectation can be evaluated using similar techniques as Equation (37).

$$
E_t[\exp(m_{t+1})] = \exp(\theta \log \delta + (\theta - 1)\kappa_0 + (\theta - 1)(\kappa_1 - 1)A_0 - (\theta - 1)A_1' Y_t) \exp(\alpha_f + \beta_f Y_t),
$$

with

$$
\begin{align*}
\alpha_f &= -\mu' \Lambda + \frac{1}{2} \Lambda' H_0 \Lambda, \\
\beta_f &= -F' \Lambda + \frac{1}{2} \Lambda' H_1 \Lambda + e_\lambda (E\nu[\exp(-\Lambda' J)] - 1).
\end{align*}
$$
Therefore the risk free rate can be expressed as

$$r_{f,t} = c^f + ((\theta - 1)A_1 - \beta_f)'Y_t,$$

where $c^f = -\theta \log \delta + (1 - \theta)(\kappa_0 + (\kappa_1 - 1)A_0) - \alpha_f$.

### C.3 Equity Risk Premium

Next, we derive the model-implied equity risk premium. Let $R_{t+1}$ be the gross return on the market. The equity risk premium is given by

$$\log(\mathbb{E}_t[R_{t+1}]) - r_{f,t} = \text{Cov}(m_{t+1}^c, r_{t+1}^c) + \log \mathbb{E}_t[\exp(m_{t+1}^j)] - \log \mathbb{E}_t[\exp(m_{t+1}^j + r_{t+1}^j)].$$

$$= B'G_tG_t' + \lambda_t(\Psi(e'_cB - 1) - \lambda_t(\Psi(e_c(B - \Lambda)) - \Psi(-e'_c\Lambda))$$

$$= ((1 - \theta)\kappa_1A_e + \gamma)\kappa_1A_e + 1)\sigma_t^2 + (1 - \theta)\kappa_1^2(A^2_\sigma^2 \phi^2_\sigma + A^2_\phi^2_\sigma)q_t$$

$$+ \lambda_t \left( (1 - \theta)\kappa_1^2A^2_\lambda^2 \phi^2_\lambda + \Psi(\kappa_1A_e + 1) - 1 - \Psi(\theta\kappa_1A_e - \gamma + 1) + \Psi((\theta - 1)\kappa_1A_e - \gamma) \right)$$

$$= \sigma^2 + q^2 + \lambda_t + \lambda^2.\sigma_t^2 + q^2 + \lambda_t + \lambda^2.$$

Here, $c^\text{ERP}_\sigma \equiv ((1 - \theta)\kappa_1A_e + \gamma)(\kappa_1A_e + 1),$ $c^\text{ERP}_q \equiv (1 - \theta)\kappa_1^2(A^2_\sigma^2 \phi^2_\sigma + A^2_\phi^2_\sigma),$ $c^\text{ERP}_\lambda \equiv (1 - \theta)\kappa_1^2A^2_\lambda^2 \phi^2_\lambda,$ and $c^\text{ERP}_j \equiv \Psi(\kappa_1A_e + 1) - 1 - \Psi(\theta\kappa_1A_e - \gamma + 1) + \Psi((\theta - 1)\kappa_1A_e - \gamma).$

### C.4 Variance Risk Premium

Before we derive the explicit form of the model-implied variance risk premium, we first show VRP has both the level difference and drift difference components. The variance of the market return is given by

$$\text{Var}(\hat{r}_{t+1} + \hat{r}_{t+2}) = \text{Var}(\hat{r}_{t+1}) + \text{Var}(\hat{r}_{t+2}) + 2\text{Cov}(\hat{r}_{t+1}, \hat{r}_{t+2}).$$
The last covariance term is zero since \( \mathbb{E}_{t+1}[\hat{r}_{t+2}] = 0 \) by assumption. VRP, therefore, can be expressed as

\[
\text{VRP}_t = \frac{\text{Var}_t(\hat{r}_{t+1}) - \text{Var}_t(\hat{r}_{t+2}) + \text{Var}_t(\hat{r}_{t+2}) - \text{Var}_t(\hat{r}_{t+2})}{2}.
\]  

Both terms I and II on the RHS of Equation (38) are non-zero. Term I captures the differences in levels of the conditional variances at time \( t \) under the physical and risk-neutral measure. As for the second term II, we apply the law of total variance and get

\[
\text{Var}_t(\hat{r}_{t+2}) = \text{Var}_t(\mathbb{E}_{t+1}[\hat{r}_{t+2}]) + \mathbb{E}_t[\text{Var}_{t+1}(\hat{r}_{t+2})] = \mathbb{E}_t[\text{Var}_{t+1}(\hat{r}_{t+2})],
\]

which is the expected change in the conditional variance. As a result, Term II can be expressed as

\[
\text{II} = \mathbb{E}_t^Q[\text{Var}_{t+1}(\hat{r}_{t+2})] - \mathbb{E}_t[\text{Var}_{t+1}(\hat{r}_{t+2})].
\]

In the definition of Drechsler and Yaron (2010), Term I is the level difference and the Term II is the drift difference.

Similar to Drechsler and Yaron (2010), the part of conditional variance coming from the Gaussian shocks, \( B'G_tG_t'B \), will cancel out in the level difference. The level difference is simply proportional to the latent jump intensity. Since the state variables have different drifts under \( P \) and \( Q \), the drift difference includes both the Gaussian- and jump-induced components. We do not repeat the analysis here.

We proceed to derive the model-implied variance risk premium. Let

\[
H_\sigma = \begin{pmatrix} \phi_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad H_\sigma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \phi_\sigma^2 & 0 & 0 \\ 0 & 0 & \phi_q^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad H_\lambda = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_\lambda^2 \end{pmatrix}.
\]

We thus have \( G_tG_t' = H_\sigma \sigma_t^2 + H_\sigma Q_t + H_\lambda \lambda_t \). Since \( \tilde{\lambda}_t = k \lambda_t, \forall t \), with \( k \equiv \mathbb{E}_\nu[\exp(-\Lambda'J)] \),

71
it holds that
\[ E_t[\lambda_{t+1}] = \mu_\lambda + \rho_\lambda \lambda_t, \]
\[ E_t^Q[\lambda_{t+1}] = k\mu_\lambda + k(-\Lambda_\lambda \phi_\lambda^2 + \rho_\lambda)\lambda_t. \]

The variance risk premium, which is defined in Equation (26), can be calculated as
\[
\text{VRP} = k\lambda_t \text{Var}^Q[J_r] - \lambda_t \text{Var}[J_r] + B'H_\sigma B(E_t^Q[\sigma^2_{t+1}] - E_t[\sigma^2_{t+1}]) + B'H_q B(E_t^Q[q_{t+1}] - E_t[q_{t+1}])
\]
\[ + B'H_\lambda B(\text{Var}_t^Q[\lambda_{t+1}] - \text{Var}_t[\lambda_{t+1}]) + E_t^Q[\lambda_{t+1}] \text{Var}^Q[J_r] - E_t[\lambda_{t+1}] \text{Var}[J_r] \]
\[
= k\lambda_t \text{Var}^Q[J_r] - \lambda_t \text{Var}[J_r] - (B'H_\sigma B\Lambda_\sigma + B'H_q B\Lambda_q \phi_q^2)q_t - B'H_\lambda B\Lambda_\lambda \phi_\lambda^2 \lambda_t
\]
\[ + \left(k\mu_\lambda + k(-\Lambda_\lambda \phi_\lambda^2 + \rho_\lambda)\lambda_t\right) \text{Var}^Q[J_r] - (\mu_\lambda + \rho_\lambda \lambda_t) \text{Var}[J_r] \]
\[ = \text{c}_t^\text{vrp} + c_q^\text{vrp} q_t + c_\lambda^\text{vrp} \lambda_t, \tag{39} \]
with the constant term \( \text{c}_t^\text{vrp} \equiv \text{Var}^Q[J_r] k\mu_\lambda - \text{Var}[J_r] \mu_\lambda \), the coefficient on \( q_t \), \( c_q^\text{vrp} \equiv -(B'H_\sigma B\Lambda_\sigma + B'H_q B\Lambda_q \phi_q^2) \), and the coefficient on \( \lambda_t \), \( c_\lambda^\text{vrp} \equiv k \text{Var}^Q[J_r] - \text{Var}[J_r] + k(-\Lambda_\lambda \phi_\lambda^2 + \rho_\lambda) \text{Var}^Q[J_r] - \rho_\lambda \text{Var}[J_r] \).

### C.5 Higher Order Risk Premium

Lastly, we derive the model-implied higher order risk premium. Define a constant \( j^M_3 \equiv E_\nu [(J_r - E_\nu[J_r])^3] \) and \( \tilde{j}^M_3 \equiv E_\nu [(J_r - E_\nu[J_r])^3] \). HRP is the difference between \( E_t^Q[M^Q_{3t+1}] \) and \( E_t[M^Q_{3t+1}] \). The details of the derivation of HRP in Equation (29) are given below
\[
\text{HRP} = \frac{1}{3} \left\{ (k\mu_\lambda - k\Lambda_\lambda \phi_\lambda^2 \lambda_t + k\rho_\lambda \lambda_t) j^M_3 - (\mu_\lambda + \rho_\lambda \lambda_t) j^M_3 \right. 
\]
\[ + \kappa_1 \phi_\lambda^2 (A_\lambda B'H_\lambda B + \text{Var}^Q(J_r)) k\lambda_t - \kappa_1 \phi_\lambda^2 (A_\lambda B'H_\lambda B + \text{Var}(J_r)) \lambda_t \right\} 
\[ = \text{c}_t^\text{hrp} + c_\lambda^\text{hrp} \lambda_t, \tag{40} \]
where the constant term \( \text{c}_t^\text{hrp} \equiv \frac{1}{3} \mu_\lambda (j^M_3 - j^M_3) \) and \( c_\lambda^\text{hrp} \equiv \frac{1}{3} \left(- (k\Lambda_\lambda \phi_\lambda^2 - k\rho_\lambda) j^M_3 - \rho_\lambda j^M_3 + \phi_\lambda^2 \kappa_1 (A_\lambda B'H_\lambda B(k - 1) + k \text{Var}^Q(J_r) - \text{Var}(J_r)) \right). \]
C.6 Calibration Details

In this section, we give details on how we calibrate the parameters of the model. The consumption process in Drechsler and Yaron (2010) can be reduced to a AR(1) process, with AR(1) coefficient given by the AR(1) coefficient of the long run risk. To adjust the time unit to half-a-month, we let $\rho_c = \sqrt{\rho_x}$, with $\rho_x$ is the AR(1) parameter of the long run risk $X_t$ in Drechsler and Yaron (2010).

Preference and volatility-related parameters are based on Bollerslev, Tauchen, and Zhou (2009). Conversion to half-a-month time unit entails

$$E[\Delta c_{t+1}] = \frac{E[\Delta c_{t+1}]}{2},$$
$$E[\sigma^2_{t+1}] = \frac{E[\sigma^2_{t+1}]}{2}, \quad \rho_{\sigma} = \sqrt{\rho_{\sigma}},$$
$$E[q_t] = \frac{E[q_t]}{2}, \quad \rho_q = \sqrt{\rho_q}, \quad \phi_q = \bar{\phi}_q,$$

where variables with a bar overhead denote the corresponding parameters in Bollerslev, Tauchen, and Zhou (2009).

According to Equation (41), $\rho_{\sigma}$, $\rho_q$, and $\phi_q$ should take values 0.988, 0.89, and $10^{-3}$, respectively, to be consistent with Bollerslev, Tauchen, and Zhou (2009). However, standard constraints need to be satisfied to ensure the existence of real-numbered solutions of second-order equations when solving for $A_0$ and $A_1$ using the method outlined in C.1. We instead set $\rho_{\sigma} = 0.9$, $\rho_q = 0.6$ and $\phi_q = 0.036 \times 10^{-3}$ in order to produce well-defined $A_0$, $A_1$ and more reasonable predictive performance.

The jump parameters are chosen to be consistent with the existing literature. Under the current parameterization, consumption rare events happen once in 20 years on average. The event intensity process is highly persistent, consistent with the small mean-reversion parameter in Wachter (2013). The jump size has a Gamma distribution, as in Drechsler and Yaron (2010). The jump size is on average negative and skewed to the left.
C.7 Setting with Three-Period-Per-Month

In the special case that there are two periods in a month, HRP is only a function on \( \lambda_t \), as shown in Section 6.2.4. Next, we show a more representative three-period-per-month case in which HRP is a function of both \( q_t \) and \( \lambda_t \). Consider the case that there are three periods in a month, we have

\[
M_3 = \mathbb{E}_t[(r_{t+1} + r_{t+2} + r_{t+3})^3]
\]

\[
= \mathbb{E}_t[r_{t+1}^3 + r_{t+2}^3 + r_{t+3}^3] + 3 \left( \text{Cov}_t(r_{t+1}, \text{Var}_{t+1}(r_{t+2})) + \text{Cov}_t(r_{t+1}, \text{Var}_{t+1}(r_{t+3})) \right)
+ \mathbb{E}_t \left( \text{Cov}_{t+1}(r_{t+2}, \text{Var}_{t+2}(r_{t+3})) \right)
\]

Similar to the two-periods-per-month case, the third moment consists of the contribution from rare events and the leverage effect. By similar arguments, the first two terms in the leverage effect

\[
\text{Cov}_t(r_{t+1}, \text{Var}_{t+1}(r_{t+2})) = \text{Cov}_t^Q(r_{t+1}, \text{Var}_{t+1}^Q(r_{t+2})),
\]

\[
\text{Cov}_t(r_{t+1}, \text{Var}_{t+1}(r_{t+3})) = \text{Cov}_t^Q(r_{t+1}, \text{Var}_{t+1}^Q(r_{t+3}))
\]

The last term, however, can be expressed as

\[
\mathbb{E}_t \left[ \text{Cov}_{t+1}(r_{t+2}, \text{Var}_{t+2}(r_{t+3})) \right]
= \kappa_1 \left( (A_\sigma B^H \sigma B + A_q B^H q_\phi \phi_q^2) \mathbb{E}_t[q_{t+1}] + (A_\lambda B^H \lambda B + \text{Var}(J_r)) \phi_\lambda^2 \mathbb{E}_t[\lambda_{t+1}] \right).
\]

Since \( \mathbb{E}_t^Q[q_{t+1}] \neq \mathbb{E}_t[q_{t+1}] \), the \( q_t \) term in the leverage effect would not be canceled after taking the difference between \( P \)- and \( Q \)-measures of \( M_3 \). Eventually, HRP is not only a function of \( \lambda_t \), but also of \( q_t \). For more than three periods cases, similar result applies.

When there are three or more periods in a month, both VRP and HRP are linear functions of \( q_t \) and \( \lambda_t \). Intuitively, since Brownian motions constitute the majority of variance, VRP should have a larger loading on the volatility-of-volatility than on event intensity. Nevertheless, volatility-of-volatility contributes to HRP only through the lever-
age premium. As a result, we expect that VRP is primarily driven by $q_t$, whereas HRP is dominated by $\lambda_t$.

As linear functions of two AR(1) processes, VRP and HRP are both ARMA(1) processes. Granger and Morris (1976) show that the persistence of such joint process depends on the two AR(1)’s own persistence and the loading coefficients. Given that $\rho_q < \rho_\lambda$, if HRP turns out to be a more persistent process than VRP, it suggests that HRP loads more on $\lambda_t$ than $q_t$, whereas VRP loads more on $q_t$ than $\lambda_t$. The empirical section (Section 3) has confirmed our conjecture. Indeed, as shown in Figure 2, HRP is more persistent than VRP empirically. The one-month AR(1) coefficient of HRP is 0.6, while that of VRP is only 0.4. Moreover, the serial correlation stays positive for as long as 15-month for HRP, while that for VRP dies out after 3 months.
<table>
<thead>
<tr>
<th></th>
<th>1-month</th>
<th>3-month</th>
<th>6-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(-0.96)</td>
<td>(-1.39)</td>
<td>(-2.05)</td>
</tr>
<tr>
<td>TVRP</td>
<td>0.31</td>
<td>0.68</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>(4.14)</td>
<td>(7.68)</td>
<td>(3.01)</td>
</tr>
<tr>
<td>VRP&lt;sub&gt;BKM&lt;/sub&gt;</td>
<td>0.35</td>
<td>0.43</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>(4.27)</td>
<td>(3.68)</td>
<td>(7.02)</td>
</tr>
<tr>
<td>HRP&lt;sub&gt;BKM&lt;/sub&gt;</td>
<td>-0.27</td>
<td>-1.65</td>
<td>-1.52</td>
</tr>
<tr>
<td></td>
<td>(-0.31)</td>
<td>(-2.87)</td>
<td>(-0.64)</td>
</tr>
<tr>
<td></td>
<td>-6.02</td>
<td>-10.14</td>
<td>-4.81</td>
</tr>
<tr>
<td></td>
<td>(-2.52)</td>
<td>(-4.24)</td>
<td>(-4.61)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-2.51)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-2.98)</td>
</tr>
<tr>
<td></td>
<td>-10.52</td>
<td>-7.41</td>
<td>-4.31</td>
</tr>
<tr>
<td></td>
<td>-11.25</td>
<td>-12.24</td>
<td>-13.66</td>
</tr>
<tr>
<td></td>
<td>-37.45</td>
<td>-19.75</td>
<td>-19.75</td>
</tr>
<tr>
<td></td>
<td>14.08</td>
<td>14.38</td>
<td>14.38</td>
</tr>
<tr>
<td></td>
<td>-7.56</td>
<td>-5.22</td>
<td>-6.67</td>
</tr>
<tr>
<td></td>
<td>12.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>9-month</th>
<th>12-month</th>
<th>24-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(-0.03)</td>
<td>(-0.13)</td>
<td>(-0.47)</td>
</tr>
<tr>
<td>TVRP</td>
<td>0.39</td>
<td>0.17</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>(1.28)</td>
<td>(0.41)</td>
<td>(0.92)</td>
</tr>
<tr>
<td>VRP&lt;sub&gt;BKM&lt;/sub&gt;</td>
<td>0.56</td>
<td>1.17</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>(1.84)</td>
<td>(3.36)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>HRP&lt;sub&gt;BKM&lt;/sub&gt;</td>
<td>-8.43</td>
<td>-12.24</td>
<td>-10.44</td>
</tr>
<tr>
<td></td>
<td>(-3.45)</td>
<td>(-3.84)</td>
<td>(-3.90)</td>
</tr>
<tr>
<td></td>
<td>-13.66</td>
<td>-19.75</td>
<td>-2.51</td>
</tr>
<tr>
<td></td>
<td>3.01</td>
<td>3.99</td>
<td>7.95</td>
</tr>
<tr>
<td></td>
<td>0.51</td>
<td>1.00</td>
<td>3.99</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>3.56</td>
<td>7.10</td>
</tr>
<tr>
<td></td>
<td>1.84</td>
<td>7.95</td>
<td></td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, $R^2$, adjusted $R^2$, and OOS $R^2$ of the predictability regressions from one to 24-month excess return on the S&P 500 index. For each horizon, we report predictive regressions results of different predictors: TVRP, VRP<sub>BKM</sub>, HRP<sub>BKM</sub>, and VRP<sub>BKM</sub> and HRP<sub>BKM</sub> jointly. VRP<sub>BKM</sub> and HRP<sub>BKM</sub> are calculated using higher moments proposed in Bakshi, Kapadia, and Madan (2003). The returns are observed monthly with the sample period ranging from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>1-month</th>
<th>3-month</th>
<th>6-month</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TVRP</td>
<td>0.31</td>
<td>0.68</td>
<td>0.61</td>
</tr>
<tr>
<td>VRP</td>
<td>0.34</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>HRP</td>
<td>-0.28</td>
<td>-0.35</td>
<td>-0.36</td>
</tr>
<tr>
<td>VRP</td>
<td>0.54</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>HRP</td>
<td>-7.80</td>
<td>-9.93</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>4.90</td>
<td>8.42</td>
<td>3.08</td>
</tr>
<tr>
<td>Adj. <strong>R^2</strong></td>
<td>4.51</td>
<td>8.04</td>
<td>2.67</td>
</tr>
<tr>
<td>OOS <strong>R^2</strong></td>
<td>3.10</td>
<td>7.76</td>
<td>-10.04</td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, $R^2$, adjusted $R^2$, and OOS $R^2$ of the predictability regressions from one to 24-month excess return on the S&P 500 index. For each horizon, we report predictive regressions results of different predictors: TVRP, VRP$^H$, HRP$^H$, and VRP$^H$ and HRP$^H$ jointly. VRP$^H$ and HRP$^H$ are calculated using Schneider and Trojani (2015) contracts. The returns are observed monthly with the sample period ranging from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.
<table>
<thead>
<tr>
<th></th>
<th>1m</th>
<th>3m</th>
<th>6m</th>
<th>9m</th>
<th>12m</th>
<th>24m</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(-1.95)</td>
<td>(-1.66)</td>
<td>(-1.33)</td>
<td>(-0.97)</td>
<td>(-0.79)</td>
<td>(-0.76)</td>
</tr>
<tr>
<td>VRP</td>
<td>0.42</td>
<td>1.01</td>
<td>1.25</td>
<td>1.13</td>
<td>1.00</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>(3.64)</td>
<td>(3.77)</td>
<td>(3.67)</td>
<td>(3.24)</td>
<td>(3.03)</td>
<td>(2.76)</td>
</tr>
<tr>
<td>JTIX</td>
<td>1.39</td>
<td>4.58</td>
<td>9.45</td>
<td>11.31</td>
<td>12.85</td>
<td>17.91</td>
</tr>
<tr>
<td></td>
<td>(2.44)</td>
<td>(4.54)</td>
<td>(3.98)</td>
<td>(3.49)</td>
<td>(3.56)</td>
<td>(2.73)</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>-0.90</td>
<td>15.50</td>
<td>10.10</td>
<td>6.40</td>
<td>4.49</td>
<td>6.51</td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(-0.74)</td>
<td>(-0.70)</td>
<td>(-0.19)</td>
<td>(0.01)</td>
<td>(0.07)</td>
<td>(-0.18)</td>
</tr>
<tr>
<td>VRP</td>
<td>0.30</td>
<td>0.73</td>
<td>0.65</td>
<td>0.45</td>
<td>0.24</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(3.81)</td>
<td>(6.93)</td>
<td>(3.01)</td>
<td>(1.33)</td>
<td>(0.52)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>RSJ</td>
<td>0.42</td>
<td>0.46</td>
<td>0.90</td>
<td>0.89</td>
<td>0.85</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(3.53)</td>
<td>(4.22)</td>
<td>(3.85)</td>
<td>(2.58)</td>
<td>(2.72)</td>
<td>(1.70)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>9.30</td>
<td>11.27</td>
<td>6.69</td>
<td>2.45</td>
<td>0.69</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>Panel C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(-0.25)</td>
<td>(-0.71)</td>
<td>(-0.59)</td>
<td>(-0.49)</td>
<td>(-0.48)</td>
<td>(-0.46)</td>
</tr>
<tr>
<td>VRP</td>
<td>0.35</td>
<td>0.77</td>
<td>0.74</td>
<td>0.52</td>
<td>0.29</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>(4.23)</td>
<td>(7.01)</td>
<td>(3.34)</td>
<td>(1.44)</td>
<td>(0.57)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>ISK</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.09</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(-0.26)</td>
<td>(-0.68)</td>
<td>(-0.87)</td>
<td>(-1.07)</td>
<td>(-0.63)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>5.05</td>
<td>9.22</td>
<td>4.10</td>
<td>1.72</td>
<td>1.13</td>
<td>0.82</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>3.16</td>
<td>9.85</td>
<td>-7.30</td>
<td>-8.87</td>
<td>-10.82</td>
<td>-8.86</td>
</tr>
<tr>
<td><strong>Panel D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(-1.27)</td>
<td>(-0.84)</td>
<td>(-0.41)</td>
<td>(-0.15)</td>
<td>(-0.06)</td>
<td>(-0.22)</td>
</tr>
<tr>
<td>VRP</td>
<td>0.34</td>
<td>0.77</td>
<td>0.77</td>
<td>0.57</td>
<td>0.37</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(4.16)</td>
<td>(6.76)</td>
<td>(3.97)</td>
<td>(1.87)</td>
<td>(0.83)</td>
<td>(1.27)</td>
</tr>
<tr>
<td>RSK</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(-0.41)</td>
<td>(-1.56)</td>
<td>(0.31)</td>
<td>(0.79)</td>
<td>(0.89)</td>
<td>(0.85)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>4.90</td>
<td>9.48</td>
<td>3.59</td>
<td>0.89</td>
<td>-0.10</td>
<td>0.31</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>4.05</td>
<td>10.99</td>
<td>-6.19</td>
<td>-8.15</td>
<td>-9.70</td>
<td>-7.95</td>
</tr>
</tbody>
</table>

The table reports the estimated regression coefficient, adjusted $R^2$, and OOS $R^2$ of the predictability regressions from one to 24-month excess return on the S&P 500 index using the variance risk premium (VRP) and other predictors. In each panel, we add one other predictor to VRP to perform predictive regressions over different horizons: JTIX in Du and Kapadia (2012) (see Equation (33)) in Panel A, realized signed jump in Guo, Sha, Wang, and Zhou (2018) (RSJ as Equation (34)) in Panel B, implied skewness in Bakshi, Kapadia, and Madan (2003) (ISK as Equation (35)) in Panel C, and realized skewness (RSK as Equation (36)) in Panel D. The returns are observed monthly with the sample period ranging from January 1996 to April 2016. Newey-West t-statistics are reported in the parenthesis.