Journal of Financial Economics 132 (2019) 150–174

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Variance risk in aggregate stock returns and time-varying return predictability

Sungjune Pyun

NUS Business School, National University of Singapore, 15 Kent Ridge Drive, Singapore 119245, Singapore

ARTICLE INFO

Article history:
Received 28 October 2016
Revised 22 November 2017
Accepted 30 January 2018
Available online 13 October 2018

JEL classification:
G11
G12
G17

Keywords:
Variance risk premium
Leverage effect
Return predictability
Beta representation
Contemporaneous beta approach

ABSTRACT

This paper introduces a new out-of-sample forecasting methodology for monthly market returns using the variance risk premium (VRP) that is both statistically and economically significant. This methodology is motivated by the 'beta representation,' which implies that the market risk premium is related to the price of variance risk by the variance risk exposure. Hence, when the slope of the contemporaneous regression of market returns on variance innovation is larger, future returns are more sharply related to the current VRP. Also, predictions are more accurate when market returns are highly correlated to variance shocks.

1. Introduction

Whether market returns are predictable using public information is of interest to both practitioners and academics. Although studies show that a number of variables can forecast future market returns, several problems have also been observed. First, predictive relationships appear to change over time, with some variables being successful in certain periods (Fama and French, 1988a) or at specific periods of the business cycle (Dangl and Halling, 2012). Second, predictors that perform well in sample often fail out of sample (Goyal and Welch, 2008; Campbell and Thompson, 2008). Lastly, return predictions typically perform poorly at shorter horizons (Fama and French, 1988a), with many well-known predictors failing to forecast returns at the horizons below six months. Statistical inference on long-horizon predictions is less reliable, raising concerns that some findings could be spurious.[1]

A recent study by Bollerslev et al. (2009) suggests that even monthly or quarterly market returns are predictable by the one-month variance risk premium (VRP), measured as the difference between option-implied variance and realized variance. They report a positive and statistically significant slope coefficient for the regression

$$R_{m,t+1} = \beta_0 + \beta_p VRP_t + \epsilon_{t+1}, \tag{1}$$

where $R_{m,t+1}$ is the leading excess market return. Theoretically, the VRP is the price of variance risk and is commonly interpreted as a proxy of time-varying aggregate risk aversion.[2] In this context, the VRP is considered to embed critical information about the moments of the stochastic

[1] See, for example, Hodrick (1992), Stambaugh (1999), Ferson et al. (2003), Ang and Bekaert (2007), and Pastor and Stambaugh (2009).

[2] See, for example, Todorov (2010), Drechsler and Yaron (2011), Bekaert et al. (2013), and Bekaert and Hoerova (2014), among others.

E-mail address: sjpyun@nus.edu.sg

https://doi.org/10.1016/j.jfineco.2018.10.002

discount factor, which are also useful in explaining variation in the market risk premium.

This paper proposes a new out-of-sample approach to monthly return predictions using the VRP that performs well both in terms of statistical and economic significance. The new methodology generates an out-of-sample R-squared of 6%–8% that is highly statistically significant, and a trading strategy produces a 0.13 gain in the annual Sharpe ratio.

The new methodology is derived from two theoretical observations. First, the one-month market risk premium should be related to the VRP by the market's exposure to variance risk. This logic follows intuitively from what is known as the "beta representation," i.e.,that the risk premium of an asset is related to the price of risk by the size of risk exposure. Empirically, this implies that the slope coefficient of the monthly predictive regression of (1) can be replaced by the market's exposure to variance risk. Empirically, the exposure ($\beta_{v,\,t}$) can be estimated by the slope of the contemporaneous regression of market returns on the unexpected changes in realized variance (RV):

$$R_{m,t} = \beta_{v,0} + \beta_v (RV_t - \mathrm{E}_{t-1}[RV_t]) + \epsilon_{o,t}. \tag{2}$$

In fact, in some respects, this estimate is far superior to the slope obtained in the traditional manner in which the predictive regression (1) is estimated directly.

The second observation is that when variance risk is responsible for a larger fraction of market risk, the VRP should explain a greater share of the market risk premium. When market risk is decomposed into two parts–a variance-related component and an unrelated component–the combination of the beta and the VRP should exactly explain the market risk premium due to the variance-related component. Empirically, this observation implies that the return predictability of the VRP would strongly depend on the size of the so-called "leverage effect," the negative relationship between market returns and variance innovation.

The traditional way of forming an out-of-sample forecast is by running the predictive regression (1) on a rolling basis for a relatively long sample. The estimated coefficients of the predictive regression are then used to form a one-step-ahead out-of-sample forecast. This traditional methodology relies on the assumption that the predictive relation remains relatively stable for an extended amount of time, so that past values of the predictive slope provide a reasonably good approximation of the predictive relationship today. However, as noted in the first paragraph, studies suggest that predictive relationships change over time. To be adaptive to time-varying predictive relation, we need a shorter estimation period. However, this may also be problematic since reducing the estimation period will increase the estimation error of the coefficients.

The new out-of-sample prediction methodology proposed in this paper rests on the close equivalence of the variance risk exposure and the predictive slope. The new approach directly uses the contemporaneous variance beta (β_v) in place of the predictive beta (β_p). As both returns and realized variances are available at the daily frequency, the contemporaneous relationship can be estimated on a monthly basis using observations from the first to the last day of the month. The size of the slope can then be multiplied by the VRP to form a return forecast for the following month.

This new methodology is potentially superior for several reasons. First, the contemporaneous regression of returns on variance innovations has a much higher R^2 compared to that of the traditional predictive regressions. A higher R^2 implies that the coefficients used for the out-of-sample predictions are estimated more accurately. Moreover, the new approach only uses the most recent month of data to determine the parameters. Hence, the proposed out-of-sample forecast methodology is applicable even when economic conditions change rapidly over time, as a result of the much shorter estimation period.

The empirical section of this paper shows that the new approach strictly outperforms the traditional way of return forecasting at the monthly horizon. In particular, the new approach predicts one-month market returns in a statistically and economically significant manner. Specifically, the traditional approach, which requires running a series of rolling predictive regressions, is unable to produce accurate forecasts of one-month returns. Across multiple VRP measures considered, some of the out-of-sample R^2s are positive (−0.8%–5.2%), but they are all far from being statistically significant. However, when we combine the VRP with the contemporaneous variance beta of the market, R^2s are always much higher (6.1%–8.4%), and the corresponding Wald statistics are always statistically significant. Finally, there is a gain of more than 0.13 (21% increase) in the annual Sharpe ratio and 4% (100% increase) in the certainty equivalent when forming a trading strategy based on the new approach.

Furthermore, the out-of-sample predictive power of the VRP depends strongly on the degree of correlation between market returns and variance innovations. When correlations are highly negative, VRP-based forecasts explain a considerable share of future market returns. On the contrary, when correlations are close to zero, market returns are essentially unpredictable by the VRP. The out-of-sample R^2s of the traditional approach are between 6.8% and 20.7% during months in which the price and variance closely move together but decrease to negative numbers when they are unrelated. When combining the VRP with the contemporaneous beta, the gap between the high and low periods slightly decreases, but the out-of-sample R^2s are always higher when correlations are more negative. This is in part because the contemporaneous beta already embeds information about the contemporaneous return-variance correlation. These results imply that the VRP provides more information about the market risk premium when returns and variance innovations move more closely together.

The in-sample results are also consistent with the hypothesis. As anticipated, the predictive beta estimated from in-sample regressions decreases in the contemporaneous variance beta. On average, a single unit decline in the contemporaneous beta leads to an approximate increase of 0.6–0.9 unit in the one-month predictive beta. In-sample predictive power also depends on the size of the correlations. The in-sample R^2 of one-month predictions ranges from 11.7%–17.9% during periods when market returns and variance innovations are highly correlated, compared to 0.4%–2.7% when the correlations are close to zero.

The ability to more directly estimate the contribution of variance risk to the market risk premium follows from three unique characteristics of the VRP. Firstly, unlike many other predictors that are merely related to some price of risk in an unknown manner, the VRP precisely measures the price of variance risk. Moreover, the underlying risk factor, namely, unexpected changes in market variance, is estimable relatively accurately using high-frequency data. Therefore, the market's exposure to variance risk is also observable, albeit with some estimation error. Finally, variance risk comprises a large part of the variation in market returns. A strong negative relation means that the observable component is likely to be an essential element of the market risk premium.

1.1. Related literature

This paper is related to at least four different areas of research. That the predictive power of the VRP might be related to the size of the leverage effect has also been hypothesized in part by Carr and Wu (2016). Bandi and Reno (2016) investigate whether there is a common shock in returns and volatility, co-jumps, that both explains a part of the market risk premium and the VRP. However, to my knowledge, this is the first paper that formally tests the relation between the leverage effect and the return predictability of the VRP.

This paper is closely related to a strand of research that studies time-varying return predictability. For example, Henkel et al. (2011) and Dangl and Halling (2012) find that the power of well-known return predictors is business cycle-dependent, and that the market is mainly dictable only during recessions.[3] Furthermore, Lettau Van Nieuwerburgh (2008) argue that there may be ts from the steady state, which makes the in-sample fficients too unstable to use for out-of-sample forecasts. Johannes et al. (2014) also show that the predictive coefficients are time-varying. Based on the idea that the predictive relation changes over time, Rapach et al. (2010) suggest combining multiple predictors to form an optimal forecast. These papers mainly study the time-varying predictability of valuation ratios (e.g., dividend yields, price-to-dividend and price-to-earnings ratios), which are known to predict returns at longer horizons. This paper focuses on the VRP, which predicts market returns over shorter horizons.

This paper also contributes to the literature that inves- he role of the price of variance risk across vari- t classes. Martin and Wagner (2016) claim that a tion of index option-implied variances is a tight und of the equity risk premium. Also, the return predictability of the variance risk premium computed from different asset classes has been investigated in a number of studies. For example, Londono (2014) and Bollerslev et al. (2014) study the predictive power of the VRP in an international context. Londono and Zhou (2017) build a variance risk premium measure from currency options and show that both the equity VRP and the currency variance risk premium are determinants of the cross-section of currency returns. Other studies include those by Wang et al. (2013) (credit default swaps) and Choi et al. (2017) (bonds). In these studies, returns from various assets are often predicted by the same underlying asset of options in which the risk premium is computed. However, the present paper suggests that this need not be the case, as long as the asset being predicted is exposed to variance risk of the U.S. stock market, it should be forecastable using the VRP computed from the Standard and Poors (S&P) 500 Index.

Because the variance beta is related to market skewness, this paper is also related to the literature on downside risk. In a recent paper, Kelly and Jiang (2014) propose a downside risk measure that predicts market returns. Feunou et al. (2017) and Bollerslev et al. (2015) suggest that it is the downside risk portion of variance risk that mostly contributes to return predictability. Carr and Wu (2016) propose an alternative measure of the VRP using an option volatility surface that better predicts market returns. Bekaert and Hoerova (2014) decompose the VIX index into two parts, one that predicts market returns and the other that proxies for financial instability. Chen et al. (2018) argue that increased financial intermediary constraints, measured using trading activities of deep out-of-the-money puts, lead to a higher risk premium. These predictors are designed to capture variation in the amount of downside market risk or variation in the pricing of that risk. The model introduced in this paper, in which downside risk is the result of a stochastic variance process, is able to account both for the amount and pricing of downside risk in a simple and coherent framework.

2. The variance risk premium and the expected market returns

In stochastic discount factor (SDF) representation, the risk premium on one-month integrated market variance, the so-called variance risk premium (VRP), can be expressed as[4]

$$VRP_T = \mathrm{Cov}_T\left(SDF_{T,T+1}, \int_T^{T+1} dV_t\right)$$
$$\approx \mathrm{E}_T^{\mathbb{Q}}\left[\int_T^{T+1} dV_t\right] - \mathrm{E}_T\left[\int_T^{T+1} dV_t\right], \tag{3}$$

where $SDF_{T,T+1}$ is the SDF over the same one-month horizon from T. The risk-neutral expectation is commonly measured using the square of the Volatility Index (VIX), available from the Chicago Board Options Exchange (CBOE). The VIX is the standard deviation of S&P 500 Index returns under the risk-neutral measure, computed using the prices of

[3] See also Garcia (2013), Chen (2009), Lustig et al. (2014), and Cujean and Hasler (2017), among others.

[4] Note the " + " sign in front of the SDF. Although the VRP can be defined as the difference between the real-world measure and the risk-neutral measure, I follow the sign convention of Bollerslev et al. (2009). Since the risk-neutral expectation of the variance is typically higher than the actual realized variance, the VRP is positive for most of the sample. The approximation comes from ignoring the effect of risk-free rates.

index options. It is then interpolated so that it matches the expectation of one-month integrated variance. Consequently, choosing the same horizon for the second component of Eq. (3) is a natural choice. The difference also has the interpretation of unit price of variance risk.

In a recent paper, Bollerslev et al. (2009) find that the VRP predicts short-term market returns. They run predictive regressions of monthly, quarterly, and semi-annual market returns ($R_{m,t+1}$) on the VRP_t,

$$R_{m,t+1} = \beta_0 + \beta_p VRP_t + \epsilon_{t+1}, \tag{4}$$

and report a positive and statistically significant β_p. The common interpretation is that the VRP is a proxy for time-varying risk aversion (Todorov, 2010; Bollerslev et al., 2011; Bekaert et al., 2013), parameter uncertainty (Bollerslev et al., 2009), or economic uncertainty (Drechsler and Yaron, 2011), so that the risk premium would change as the moments of the SDF vary.

Besides being a short-term return predictor, the VRP is unique for several other reasons. One is that it is actually a price of risk, rather than a variable that merely encodes information about risk prices (e.g., the dividend yield) in an unknown manner. Second, the factor on which it is based, namely, variance innovations, can be observed with a tolerable amount of estimation error. Most importantly, those variance innovations are highly correlated with market returns, which means that the market has substantial exposure to variance risk.

Prior research often ignores an important aspect of the equity market, namely, that the market price and variance tend to move in the opposite direction. One well-known explanation, known as the "leverage" effect (Black, 1976; Christie, 1982), hypothesizes that an adverse shock in the market causes the overall leverage to increase, leading to higher volatility. Hence, the level of variance and price are negatively related. A more popular explanation, known as "volatility feedback," is that risk-averse investors require a higher premium for being in a high-volatility state. Therefore, investors demand more compensation in the future when market variance increases unexpectedly. Since a higher risk premium implies a lower value today, the market price must drop when variance increases (Pindyck, 1984; French et al., 1987; Bollerslev et al., 2006).

This negative relationship implies that the market portfolio is subject to variance risk, which naturally suggests that VRP is the premium on an important source of aggregate variation in the stock market that also affects its required returns. Moreover, since the VRP can be measured relatively accurately, a fraction of the market risk premium can be inferred from the VRP, essentially in real time. The following simple model demonstrates this relationship in detail.

2.1. A simple model

To build some intuition, consider a stochastic volatility model in which the correlation (ρ_t) between market returns and changes in market variance is assumed to be time-varying.[5] When S_t is the price of the aggregate market portfolio, approximately represented by the S&P 500 Index, and V_t is its variance, we have

$$\frac{dS_t}{S_t} = \mu_t dt + \sqrt{V_t}\left(\rho_t dW_t^v + \sqrt{1-\rho_t^2}\, dW_t^o\right) \tag{5}$$

$$dV_t = \theta_t dt + \sigma_v dW_t^v. \tag{6}$$

By construction, the two Brownian motions dW_t^v and dW_t^o are independent. These processes assume that the return and variance follow a bivariate Gaussian process with a negative correlation. The drifts are not specified but are assumed to be time-varying. The volatility of the variance is assumed to be constant but could also be time-varying. Solving the first equation in terms of variance innovations and dW_t^o yields

$$\frac{dS_t}{S_t} = \mu_t dt + \rho_t \frac{\sqrt{V_t}}{\sigma_v}(dV_t - \theta_t dt) + \sqrt{(1-\rho_t^2)V_t}\, dW_t^o. \tag{7}$$

This two-factor structure indicates that market movements can be decomposed into two parts. First, market prices can vary as market variance moves. For example, in the intertemporal model of Merton (1973), an unexpected increase in the market variance must directly lead to lower returns. The second part reflects price movements due to all other reasons. They may include shocks from the real economy perhaps to production or consumption, which can be orthogonal to market variance shocks. By rotational indeterminacy, an orthogonal decomposition of real economic shocks is always feasible. For simplicity, I, therefore, refer to the second term as the 'orthogonal' component and the premium associated with this component as the orthogonal premium.

The SDF representation can be used to match the one-month VRP with the market risk premium for the same interval. The monthly market risk premium can be expressed as

$$\begin{aligned}
&\mathrm{Cov}_T\left(-SDF_{T,T+1}, \int_T^{T+1} \frac{dS_t}{S_t}\right) \\
&\quad = -\rho_T \frac{V_T}{\sigma_v}\, \mathrm{Cov}_T\left(SDF_{T,T+1}, \int_T^{T+1} dV_t\right) \\
&\qquad - \sqrt{(1-\rho_T^2)V_T}\, \mathrm{Cov}_T\left(SDF_{T,T+1}, \int_T^{T+1} dW_t^o\right).
\end{aligned} \tag{8}$$

Eq. (8) is the key to understanding my approach and represents the relationship between the two risk premia in continuous time. This equation suggests that the market risk premium can be decomposed into a linear combination of two prices of risk. The first part governs how the one-month VRP relates to the market risk premium. Notably, the size of the slope that connects the VRP to the market risk premium ($-\rho_t \frac{V_t}{\sigma_v}$) is the negative of the market's exposure to variance risk. When we run a simple linear regression of market returns on its contemporaneous variance shocks, this term is *exactly* the slope of this regression. Thus, the relation between returns and unexpected changes in variance determines the slope that connects the VRP and *future* market returns.

In fact, the slope measures how the market responds to unexpected changes in market variance. From the perspective of an investor who holds the market portfolio and

[5] Note that ρ_t is typically negative.

wants to reduce exposure to variance risk, the first component of Eq. (8) represents the market risk premium due to a part that can be hedged using a variance swap. A variance swap exchanges future realized variance for a notional amount. Carr and Wu (2009) show that the risk-neutral expectation of variance is the notional amount of the swap. Therefore, the VRP is essentially the expected unit cost of variance risk, and the contemporaneous variance beta is the number of swap contracts required to hedge the market portfolio against variance movement. The combination is what the investor needs to pay to hedge variance risk.

The second term of Eq. (8) represents how the price of the orthogonal component relates to the market premium. The orthogonal premium could potentially affect how the VRP and the market risk premium are related because the VRP and the orthogonal premium could also be related. For example, an increase in aggregate risk aversion could both affect the VRP and the orthogonal premium simultaneously. Therefore, depending on the degree to which the two premia are linked, it is possible that the orthogonal premium modifies how the VRP relates to the market risk premium.

There are at least two reasons to believe that orthogonal risk is largely unrelated to the VRP. First, while the predictive power of the VRP decreases as the forecast horizon increases, the opposite is true for other well-known predictors, such as the dividend yield (Fama and French, 1988a), earnings-to-price ratio (Campbell and Shiller, 1988), term spread (Campbell, 1987), and *cay* (Lettau and Ludvigsen, 2001). Moreover, these predictors tend to perform well during recessions. For example, Rapach et al. (2010), Henkel et al. (2011), and Dangl and Halling (2012) demonstrate that the predictive power is strong only during recessions. As will be shown in the following section, these periods do not coincide with periods in which there is a strong negative relation between market returns and market variance, which is when the VRP has its strongest predictive power. This evidence suggests that there is at least a different source of risk that is not too much related to market variance shocks.

Although not entirely realistic, there are at least two circumstances where the linear relationship between the VRP and the market premium would be exact. Under these assumptions, the slope that governs the relationship between the premia is exactly the slope that connects returns to variance innovations. The first case is when the orthogonal component is unpriced. The other is when the orthogonal premium is uncorrelated with the VRP.

2.2. Empirical implications

The remainder of this paper thoroughly discusses several important aspects of the simple model presented. First, the slope that determines the relation between the short-term market risk premium and the risk premium on market variance is largely determined by the amount of variance risk present in the market portfolio. We can run a contemporaneous regression of the daily excess market returns on the unexpected change in realized variance (RV) over a fixed interval as

$$R_{m,t} = \beta_{v,0} + \beta_v (RV_t - \mathrm{E}_{t-1}[RV_t]) + \epsilon_{o,t}. \tag{9}$$

The slope of this regression measures how much the market reacts to unexpected changes in market variance. This is also the coefficient on the VRP in (8). The equation that relates the VRP to the market risk premium can then be represented as

$$\mathrm{E}_T[R_{m,T+1}] = -\beta_v VRP_T + O_T, \tag{10}$$

where O_T is the premium due to the orthogonal component.[6] The orthogonal premium is theoretically equivalent to $\sqrt{V_T(1-\rho_T^2)}\ \mathrm{Cov}_T(SDF_{T,T+1}, \int_T^{T+1} dW_t^o)$ from (8).

Second, the equation that describes the relation between the VRP and the market risk premium suggests that market returns are more accurately predictable when the index and the variance of returns move closely together. The model indicates that the proportion of the total market variation related to variance risk is ρ_t^2. If the orthogonal premium is unpriced or unrelated to the VRP, the orthogonal premium will appear as noise in a predictive regression in which the VRP is the sole predictor. If the premium is priced and related to the VRP, this premium will bias the predictive beta. In either case, as the contemporaneous correlation (ρ_t) gets closer to zero, predictions will become less accurate. On the other hand, when correlations are close to −1, the VRP should almost entirely identify the market risk premium. One way to understand this is to consider variance swaps discussed earlier. A variance swap can perfectly hedge the market portfolio. Under no arbitrage, a perfectly hedged position should not generate anything more than the risk-free rate. Conclusively, the predictive R^2 should depend on the size of the correlation between market returns and variance shocks.

These two relations are extremely useful when forecasting market returns out of sample. As discussed above, accurate predictions of market returns are extremely hard. In-sample R^2s rarely exceed 5% for most common predictors (Goyal and Welch, 2008) at the annual horizon. A low R^2 also implies that the parameter estimates are likely to be more inaccurate, which induces poor out-of-sample forecasts. To compensate for the high percentage of noise in returns data, an extended estimation sample is required. However, this is only advisable when the predictive relation remains constant.

However, recent research suggests that the predictive relationship does not remain constant over time. As noted a[illegible]aluation ratios such as the dividend yields, t[illegible]e power is higher during recessions. Also, J[illegible] al. (2014) argue that the parameters that g[illegible]predictive relation are time-varying. As the t[illegible]ay of using predictive regressions assumes a constant predictive relationship, forecasts can be particularly misleading when the relation evolves rapidly over

[6] The relationship between the one-month market risk premium and the VRP can also be roughly observed by taking expectations on both sides of the contemporaneous regression with respect to the risk-neutral measure and the real-world measure and subtracting one from the other. The one-month risk premium is exactly the sum of the premium on the two components.

time, such as when there is a structural break. The econometrician may attempt to address parameter instability by using a short estimation window. However, when the first-stage estimation period is too short, as mentioned, the coefficients of the first-stage predictive regression will be imprecise.

There is an alternative approach that can be used specifically for the VRP, which I refer to as the "contemporaneous beta approach." The method is based on the close relationship between the predictive and contemporaneous betas and implemented by using the beta of the contemporaneous regression in place of the predictive beta to form the out-of-sample forecast.

There are two reasons why the contemporaneous beta could potentially be a much more accurate estimate than the rolling-window predictive beta. First of all, the contemporaneous relation between returns and changes in variance is much stronger than the predictive relationship between the VRP and future returns. While the predictive R^2s [illegible]ly exceed 5%, the average of the R^2s of the contemporaneous regression is slightly above 15%.

Furthermore, both returns and estimates of realized variance are available at the daily frequency. Being able to use data at a higher frequency implies that there are more data, which no longer necessitates relying on an extended estimation period. Using a short estimation period, such as a month, might be enough to generate a slope coefficient that is sufficiently accurate to form an out-of-sample forecast. Depending on what is assumed for the orthogonal premium and variance forecast, it is essentially possible to get an estimate of the monthly equity premium using a single month of data. Hence, this new approach can be used even when the predictive relation changes rapidly over time.

Finally, in principle, a similar logic should also apply to asset classes other than the equity index. It is not necessary to use a VRP that is based on the asset whose returns one is trying to predict. As long as those returns are correlated with changes in S&P 500 Index volatility, the VRP should have some predictive power. Assets that are highly correlated with variance innovations should be predictable with higher accuracy, and those that relate to changes in market variance with a higher beta should be predictable using the VRP with a higher slope.

The following sections examine supporting evidence that shows that the contemporaneous and predictive relations for market returns are in fact closely connected.

3. Data and estimation

This section discusses how the VRP, the contemporaneous betas (denoted by $\hat{\beta}_v$), and the correlations (denoted by $\hat{\rho}$) are measured from daily market returns and RVs.

3.1. Forecasting variance

To estimate the VRP, selecting a good variance forecast model is important. As Bekaert and Hoerova (2014) argue, the VRP's ability to predict returns may depend on the particular model used to compute the real-world expectation component. I use intraday, high-frequency, return-based RV to model the forecasts. It is known that high-frequency RV models have advantages over standard generalized autoregressive conditional heteroscedasticity (GARCH) or stochastic volatility (SV) models, which typically rely on daily returns. First, traditional GARCH or SV models are somewhat difficult to estimate. Distributional assumptions are required for either model. Moreover, RV-based models are known to outperform standard GARCH or SV models when forecasting variance. This outperformance is partly possible because high-frequency data allows one to measure the latent variance process more accurately. Finally, using RV allows to fit complex multivariate models that capture the long memory feature of the latent variance process.[7]

The high-frequency intraday trading data for the S&P 500 Index are obtained from Tickdata. The data are available from 1983, but this paper only requires the data between 1989 and 2016 since the first component used to estimate the VRP, the VIX^2, is only available from 1990. To estimate the second component, RV is computed by first calculating squared log returns from the last tick of each five-minute interval. A subsampling scheme at one-minute intervals (Zhang et al., 2005) is used to reduce microstructure noise. Hansen and Lunde (2006), for example, study the impact of subsampling and note that, theoretically, it is always beneficial in reducing microstructure noise. I rescale the RVs to the monthly level so that they match the variance of a month.

The constructed RV series is then used to compute the variance forecasts. Corsi (2009) proposes a Heterogeneous Autoregressive Realized Volatility (HAR-RV) model. The model assumes that the predicted value of volatility is linear in its autoregressive components–daily, weekly, and monthly realized volatility. By distinguishing a long-run monthly component from the short-run daily component, the model performs well in capturing short-term variation in the volatility process together with the long memory feature of volatility.

I use a variation of Corsi's model and forecast the RV instead of the volatility. The market variance of day $\tau + k$, for any $k \geq 1$ can be forecasted on day τ using past values of RVs by running the following regression:

$$RV_{\tau+k} = a_0 + a_d RV_\tau + a_w \sum_{j=0}^{4} RV_{\tau-j} + a_m \sum_{j=0}^{21} RV_{\tau-j} + e_{\tau+k}. \tag{11}$$

To avoid confusion, I use τ for variables that are in daily frequency and t subscript for a variable that is in monthly frequency.

The one-day RV forecast $(\widehat{RV}_{\tau+1|\tau})$ can then be constructed using the loadings on the daily, weekly, and monthly components. The forecast of monthly variance at

[7] Under several conditions, Andersen et al. (2001) show that the RV converges in probability to the true variance. Andersen and Bollerslev (1997) indicate that variance estimates based on intraday returns provide information about long-run volatility dependencies. For advantages using high-frequency-based RV, see Andersen et al. (2003). Also see Andersen et al. (2007), Chen and Ghysels (2011), and Busch et al. (2011), among others for details about the performances and extensions of the heterogeneous autoregressive-type models.

day τ ($\widehat{RV}_{\tau+1,\tau+22|\tau}$) is estimated by averaging the 22 daily forecasts ($k = 1, \ldots, 22$).[8] The forecasts are estimated using daily observations on a 12-month rolling window to account for the possibility that the forecast relation changes over time.

3.2. Estimation of the VRP

The VRP is measured by taking the difference between the square of VIX and the monthly forecast of RV. While the end-of-month values of the VIX squared are typically used for the first component, in the literature, the second component is estimated by using a forecast model on the RV computed by summing up daily observations over the entire month.[9] This approach suffers from a timing mismatch, especially when the variance is trending during the most recent month. While VIX^2 reflects changes in market variance during a month, the above forecast of monthly variance does not account for possible changes in market variance during the most recent month. Even a small trend in market variance may impact the predictability of the VRP, especially, if these variance trends are related to the degree of the negative correlation between market returns and variance shocks.

Two direct solutions are implementable to minimize any volatility trends affecting the VRP. One can either choose to average both of these values over the month or use the end-of-month values for both components. Averaging over daily values reduces possible estimation error or the influence of a single observation, but it may not reflect the most up-to-date information. On the other hand, using the end-of-month value reflects more recent information, but is more likely to be subject to estimation error. Following the literature, I estimate the VRP as the difference between VIX^2 and the 22-day cumulative forecast of daily realized variance. However, to deal with the mismatch, I either average the daily observations or take the end-of-month values. These two measures are parametric and denoted by $VRP_{\overline{P}}$ and VRP_{P^E}, respectively.

$$VRP_{\overline{P},t} = \sum_{\tau \in t} \left(\frac{VIX_\tau^2}{252} - \frac{\widehat{RV}_{\tau+1,\tau+22|\tau}}{22} \right) \tag{12}$$

$$VRP_{P^E,t} = \frac{VIX_{m(t)}^2}{12} - \widehat{RV}_{m(t)+1,m(t)+22|m(t)}, \tag{13}$$

where $m(t)$ is the last trading day of month t. Both of the variance components are rescaled to match the one-month variance. While this choice affects the predictive performance when running a simple predictive regression, I show that the key results of this paper hold regardless of the timing of the measurement.

To ensure that the empirical results are not dependent on a particular model, I supplement these two measures with a third nonparametric one. Denoted by VRP_N, the nonparametric VRP is the difference between the scaled VIX^2 and the historical RV, both averaged over the entire month. This VRP is similar to the one used by Bollerslev et al. (2009), who take the difference between the end-of-month value of the implied variance and the monthly RV. Here, I use the monthly average for both to avoid the possibility of volatility trends affecting the VRP.

Fig. 1 compares the time series of the three VRPs used in this paper. They are highly correlated to each other, but the parametric one is more persistent than the nonparametric VRP, and between the two parametric measures, the VRP using the monthly average tends to be more persistent than the one observed at the end of the month. Also acknowledged by previous studies (e.g., Bali and Zhou, 2016), there is a negative spike during the Financial Crisis of September 2008. This is because RV was unexpectedly high at this time and reverted to its original level swiftly. The VIX did not increase as much during that period because, presumably, part of the spike was regarded as a jump in the index. Therefore, the forecast models did not appear to have captured the strength of the mean reversion that was observed.

3.3. Estimation of the contemporaneous betas and correlations

The daily innovation of market variance is calculated by computing the unexpected changes in RV scaled so that it matches the one-month interval. Then, the monthly contemporaneous beta is estimated from the regression of market returns on variance innovations, using only observations that belong to that particular month.

$$R_{m,\tau} = \beta_{v,0,t} + \beta_{v,t}(RV_\tau - \widehat{RV}_{\tau|\tau-1}) + \epsilon_\tau. \tag{14}$$

The choice of the estimation window follows that of Ang et al. (2006) and Chang et al. (2013). They also use a single month of data to estimate the variance betas for individual stocks. One concern is that the variance of the error term ϵ_τ is likely to be correlated with the explanatory variable. To deal with possible heteroscedasticity, I consider weighted least squares (WLS) in addition to ordinary least squares (OLS). To distinguish them from each other, I use $\hat{\beta}_v$ to denote the OLS estimates and $\hat{\beta}_{v,WLS}$ to denote the WLS estimates.

The contemporaneous correlation ($\hat{\rho}_t$) is the correlation between the two variables in the above equation. The correlations are closely connected to the betas because they are transformations of each other.

$$\hat{\rho}_t = \hat{\beta}_{v,t} \times \frac{\hat{\sigma}_t(RV_\tau - \widehat{RV}_{\tau|\tau-1})}{\hat{\sigma}_t(R_{m,\tau})}. \tag{15}$$

Because each regression is based on observations from a single month, the monthly series of betas and correlations are estimated from non-overlapping samples.

The time series of the betas is provided in Fig. 2 and that of the correlations is in Fig. 3. The dotted line shows the time series of the one-month estimates. The three-month estimates in solid lines supplement the one-month

[8] Note that this is equivalent to running one predictive regression with the sum of the 22 dependent variables, but the difference is what can be used in a single regression. On day τ, the method I use uses all past RVs up until day $\tau - 1$, while having the sum of 22 RVs can only use all data up to $\tau - 22$.

[9] See, for example, Bollerslev et al. (2009), Bekaert and Hoerova (2014), and Gonzalez-Urteaga and Rubio (2016), among others.

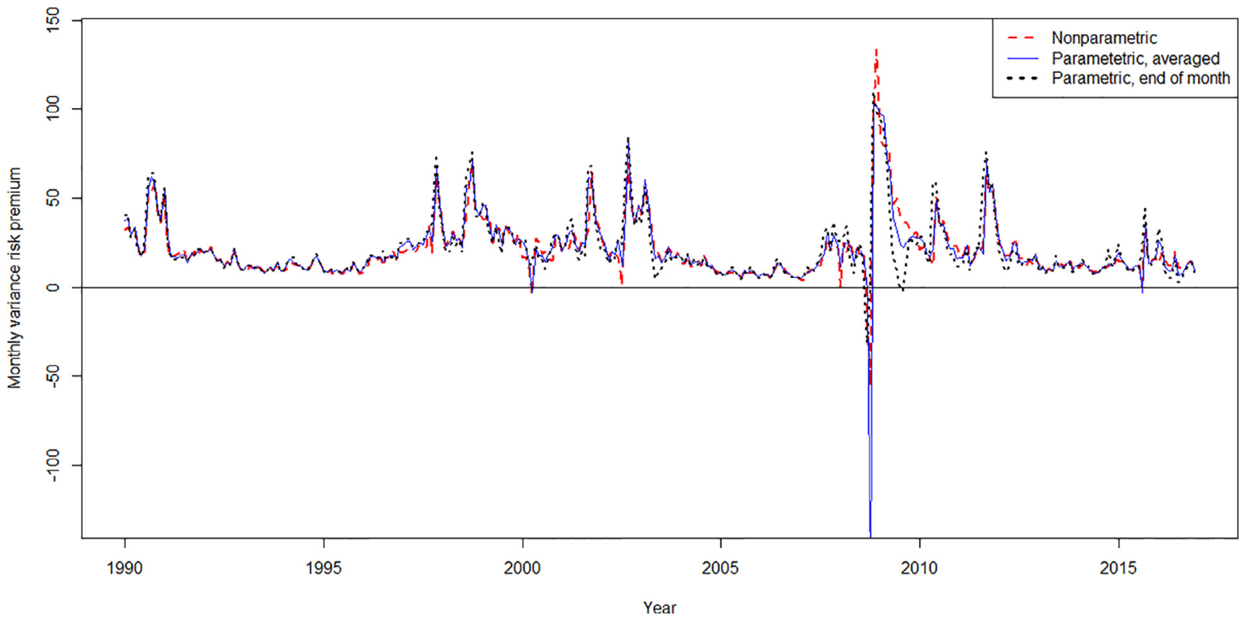


Fig. 1. Time series of the variance risk premium (VRP). This figure compares the three VRP measures considered in this paper. VRP_N is nonparametric, measured as the difference between the monthly average of $VIX^2/12$ and the monthly realized variance. Two other measures are model-based and estimated as the difference between $VIX^2/12$ and the one-step-ahead forecast of the realized variance. A variant of the HAR-RV model of Corsi (2009) is used to predict the variance. It is either averaged over the month ($VRP_{\overline{P}}$) or estimated at the end of the month (VRP_{P_E}).

estimates.[10] Also in Fig. 2, both the WLS (top) and OLS (bottom) betas are provided as a separate figure.

As can be observed from these plots, both the betas and correlations are quite volatile. Several other remarks are worth noting. First, the contemporaneous correlations were more negative during the second half of the sample (2004–2016) with a coefficient of −0.308 as opposed to −0.229 during the first half of the sample. The difference in the size of the negative correlation indicates that the VRP may be more effective as a return predictor during the second half of the sample. These are also times when the fluctuations in the betas are better observed. Third, especially for the post-2000 period, the leverage effect becomes weaker several months after negative market shocks. These shocks include the 1997 Asian Crisis, dot-com burst of 2001–2002, the Financial Crisis of 2008, and the Shanghai market crash of 2015. However, some of the positive movements (e.g., 2004–2005 and 2013) in betas do not follow negative market shocks. These are more likely to be times when the market rebounded following a negative shock, and during these times, the VRP did not provide high-quality information about the short-term market risk premium.

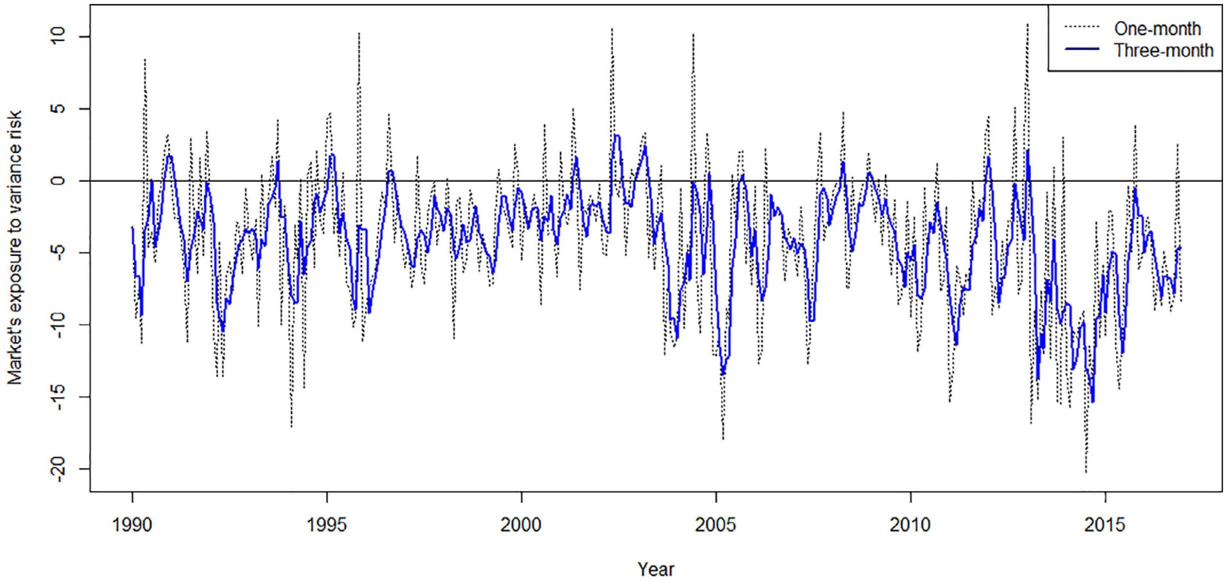
3.4. Summary statistics

Table 1 provides summary statistics for the key variables of interest. These variables include RV, option-implied variance, three measures of the VRP, the contemporaneous variance betas, and the contemporaneous correlation. The study period is from January of 1990 to December of 2016, which is restricted by the availability of the VIX. There are a total of 324 months, 37 of which the National Bureau of Economic Research (NBER) classifies as recessions. There are three recession periods, one in 1990–1991, another in 2001, and the last one in 2008–2009.

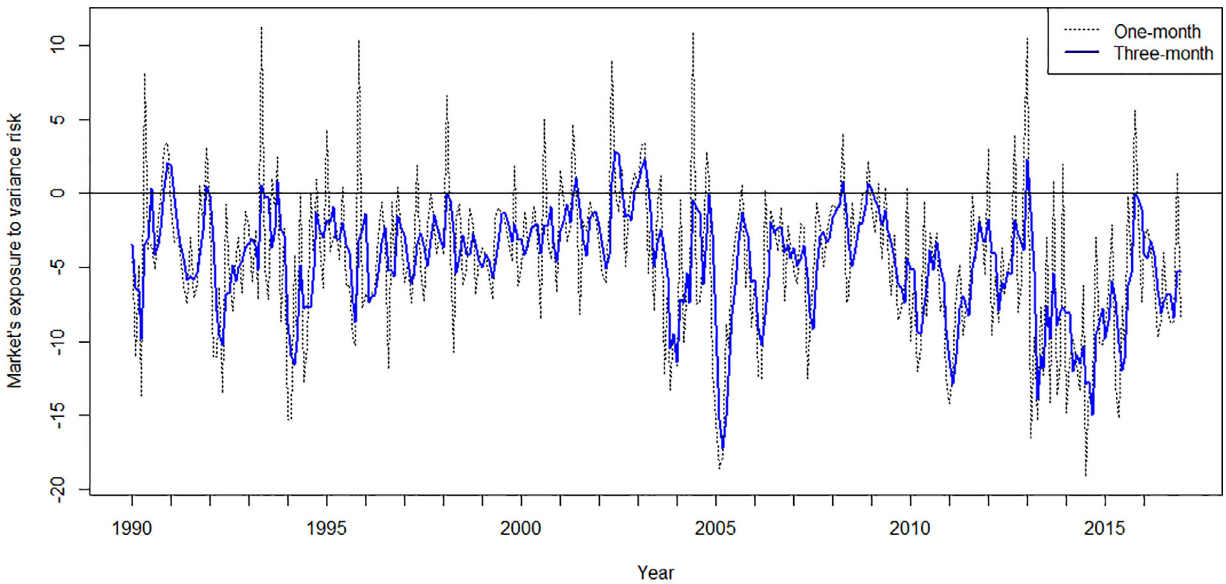
The first two columns of the table summarize the means and standard deviations over the entire sample. Then, the next columns summarize the statistics for several subsamples. I first divide the sample into two sub-periods, one in which the contemporaneous correlation is greater than and the other in which it is less than the median of the entire sample. The table only reports the statistics for the "greater" periods. This classification is useful for testing whether the key results of this paper are results of a quadratic predictive relation. If there is a quadratic predictive relation, a higher predictive slope would be observed during times with a high VRP. The question is whether the VRP is higher when the contemporaneous correlation between returns and variance shocks is more negative. As the summary statistics suggest, the levels of the variance and the VRPs are similar regardless of whether the correlation is higher or lower than the time-series median. Hence, the time-varying return predictability is unlikely to be driven by a quadratic predictive relation.

Recent studies suggest that market returns can be predicted better during recession periods. For example, Pesaran and Timmermann (1995), Henkel et al. (2011), and Dangl and Halling (2012) argue that the performance of traditional predictors, such as the dividend yield, is strong only during recessions. The set of predictors they consider does not include the VRP. To rule out the possibility that the contemporaneous correlation is not merely a proxy for

[10] The three-month betas are estimated three-month moving averages, and the three-month estimates are computed by taking the moving average of each of the components in Eq. (15) separately.



(a) Weighted least squares



(b) Ordinary least squares

Fig. 2. Time series of the contemporaneous betas. This time-series plot shows the time-variation in the monthly contemporaneous betas. The contemporaneous beta is the slope of the regression of market returns on variance innovations.

$$R_{m,\tau} = \beta_{0,t} + \beta_{v,t}(RV_\tau - E_{\tau-1}[RV_\tau]) + \epsilon_\tau. \quad (28)$$

This regression is estimated using all observations $\tau = \{\tau_1, \ldots, \tau_{m(t)}\}$ that belong to month t. The regression is estimated using (a) weighted least squares or (b) ordinary least squares. The solid line is the three-month moving average of these estimates.

business cycles, I ask whether the correlations are more negative during recession periods.

The next two columns of Table 1 show descriptive statistics of the variables during NBER recession periods. The statistics show that contemporaneous correlations are not apparently more negative during recessions. Two implications are worth noting. First, the statistics show that the findings of this paper are not implied by the work discussed above, in that I am not showing a pattern in predictability that is related to business cycles. As will be seen in the next several sections, the predictability of the VRP is higher when the leverage effect is stronger. However, the statistics suggest that these are not necessarily the periods where traditional predictors tend to perform

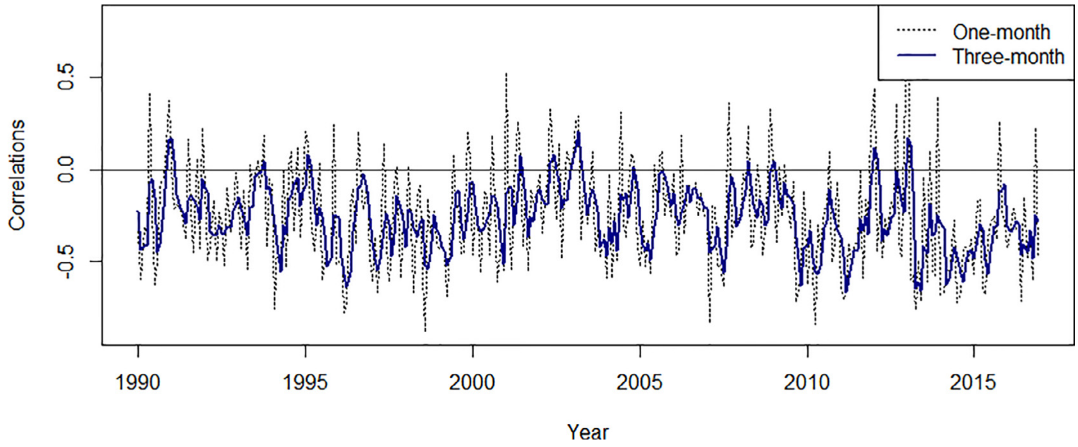


Fig. 3. Time series of the contemporaneous correlations. This time-series plot shows the time-variation in the monthly contemporaneous correlations between daily returns and variance innovations. Then, these correlations are computed using all daily observations that belong to the month. Variance innovations are defined as the difference between realized variance and the lagged one-step-ahead variance forecast of the HAR-RV model. The solid line is computed using three months of data.

Table 1
Summary statistics.

Panel A of this table summarizes the means, standard deviations, and the first-order autocorrelations for the main variables of interest during the sample period of 1990–2016. The realized variance (RV) is the sum of the square of five-minute market returns from the first to the last day of the month. The implied variance (IV) is the monthly mean of the volatility index (VIX) squared divided by 12. The variance risk premium (VRP) is estimated using three different methods. VRP_N is the monthly average of the difference between IV and RV. $\text{VRP}_{\overline{P}}$ is the difference between IV and the sum of the 22-day RV forecasts of the month. VRP_{P^E} is the same parametric measure but estimated at the end of the month. These variables are multiplied by 10,000. The contemporaneous beta of month t is the slope of the regression

$$R_{m,\tau} = \beta_{0,t} + \beta_{v,t}(RV_\tau - E_{\tau-1}[RV_\tau]) + \epsilon_\tau, \tag{31}$$

where $R_{m,\tau}$ is the market excess return of day τ, estimated using all observations during the month. $\hat{\beta}_{v,WLS,t}$ is the estimate from weighted least squares. The contemporaneous correlation ($\hat{\rho}_t$) is the correlation between daily market returns and variance innovations. Panel B summarizes the correlations between the beta/correlation estimates and other variables of interest. VIX trend is the difference between the end-of-month level and the monthly average of VIX. SKEW is the CBOE SKEW index, and tail risk is from Kelly and Jiang (2014).

Panel A: Summary statistics

			$\hat{\rho} \le$ median		NBER recession		
	Mean	StDev	Mean	StDev	Mean	StDev	Autocorr.
RV (monthly)	16.43	29.51	16.67	34.68	47.00	72.52	0.643
Implied variance (monthly)	19.71	7.49	19.89	7.34	29.02	10.30	0.840
VRP_N	20.96	16.74	19.01	14.63	38.13	32.58	0.696
$\text{VRP}_{\overline{P}}$	21.74	17.86	21.43	16.38	39.76	31.01	0.764
VRP_{P^E}	21.30	18.77	21.86	19.98	31.81	29.04	0.601
$\hat{\beta}_v$	−4.19	5.10	−7.72	3.92	−1.49	3.05	0.200
$\hat{\beta}_{v,WLS}$	−4.56	5.08	−7.72	3.81	−1.60	3.01	0.203
$\hat{\rho}$	−0.259	0.281	−0.486	0.132	−0.130	0.226	0.124
Number of months	324		162		37		

Panel B: The leverage effect and correlations

	$\hat{\beta}_{v,t+1}$	$\hat{\rho}_{t+1}$
Contemporaneous returns ($R_{M,t+1}$)	0.120	0.280
Lagged annual returns ($\sum_{k=0}^{11} R_{M,t-k}$)	−0.261	−0.287
RV_t	0.196	0.142
VIX_t	0.297	0.165
VIX trend_t	−0.121	−0.259
SKEW_t	−0.300	−0.222
Tail risk_t	−0.128	−0.131
$\text{VRP}_{N,\,t}$	0.199	0.152
$\text{VRP}_{\overline{P},t}$	0.274	0.214
$\text{VRP}_{P^E,t}$	0.220	0.128

well. Second, as mentioned earlier, the predictability of the VRP is stronger for short-horizon returns while other common predictors are stronger over longer horizons. These two facts suggest that the empirical findings of this paper are somewhat independent of earlier findings that return predictability tends to be stronger during recessions. Also, they strengthen the hypothesis that the market risk premium can be decomposed at least by two parts discernible from each other.

The last column of the table summarizes the first-order serial correlations of these variables. Overall, the moderate level of these serial correlations suggests that these variables are stationary. The autocorrelations of the contemporaneous betas are slightly higher than 0.2, and that of the contemporaneous correlation is slightly above 0.1. By construction, the autocorrelations of the overlapping three-month OLS betas and correlations are higher, with 0.79 and 0.66, respectively (not reported in the table).

To gain some insight into what drives the negative relationship between market returns and variance innovations, Panel B of the table summarizes the contemporaneous and predictive correlation between the leverage estimates and several other variables of interest. I compare the estimates with monthly contemporaneous market returns and annual lagged returns, the variance of the market, and several measures of downside risk.

First, the table suggests that the leverage effect tends to be stronger following positive shocks in the market. This result is consistent with the earlier argument that the VRP tends to be less informative during bad economic times. Also, both betas and correlations are more negative when VIX is low. The implications are consistent with Johnson (2017), who suggests that the VRP tends to be a noisy measure of market risk premium during high volatility periods. Generally, VRP should be most informative about the expected market returns during low volatility and high valuation times.

Second, the leverage effect is stronger when there is a positive trend in market variance. That is, the VRP is likely to be more informative when variance increases. The next line of the panel directly shows this. Here, the VIX trend is defined as the difference between the VIX at the end of the month and the average over the month. Overall, these two pieces of evidence confirm that the VRP tends to be least informative when the market rebounds following a negative shock.

I also compare these estimates with several proxies for downside risk. Downside risk may be related to the leverage effect since when there is more downside risk, the market may be more sensitive to small changes in the level of variance. I consider the SKEW index of CBOE, the tail risk of Kelly and Jiang (2014), and the three VRPs. The table suggests that the leverage effect is stronger when the market is more negatively skewed, but when the VRPs are low. Again, this is consistent with the argument that the VRP is more informative during less volatile periods.

4. Out-of-sample predictions

This section documents two novel and striking findings. First, the beta that explains the predictive relationship is close to the negative of the contemporaneous beta. They are, in fact, so close that the contemporaneous beta can be directly used in place of the predictive beta for the out-of-sample (OOS) forecast. Second, predictions perform better when the contemporaneous correlation between market returns and variance innovations is more negative. The first part of this section provides the main result of this paper, OOS predictions. The performance of possible trading strategies follows. The next section discusses in-sample prediction results.

I focus on one-month returns because the VRP is the premium over a one-month horizon. Therefore, evaluating market returns at the same horizon is a natural choice, as suggested by the model. There are also other reasons for doing so. The contemporaneous relation between return and variance varies rapidly over time, as shown in Fig. 2. Therefore, the contemporaneous relationship estimated based on past data may not be valid over a much longer horizon.

4.1. Out-of-sample predictions

The traditional approach to providing OOS forecasts of time $T+1$ returns consists of two stages. First, I run a predictive regression using the past k months of historical data (from time $T-k+1$ to T) as

$$R_{m,t} = \beta_0 + \beta_p VRP_{t-1} + \epsilon_t. \tag{16}$$

We typically use the coefficient estimated at time T to forecast returns at time $T+1$. Then, the one-step-ahead predicted value of the excess market returns ($\widehat{R}_{m,T+1|T}$) is given as $\hat{\beta}_{0,T} + \hat{\beta}_{p,T} VRP_T$.

The next step is to evaluate the OOS predictive performance, for example, using the OOS-R^2. To do so, Goyal and Welch (2008) and Campbell and Thompson (2008), among others, compute the OOS-R^2, defined as

$$1 - \frac{\sum_t (\widehat{R}_{m,t+1|t} - R_{m,t+1})^2}{\sum_t (\bar{R}_{m,t} - R_{m,t+1})^2}, \tag{17}$$

where $\bar{R}_{m,t}$ is the historical average of the market returns evaluated at time t. Finally, we compute a test statistic, for example, a Wald statistic, to test the significance of the predictor. Diebold and Mariano (1995) provide a formal test for such OOS prediction errors. Giacomini and White (2006) extend the OOS test and propose a Wald test that is valid for testing nested models. The Wald statistic is given as

$$W = T\left(T^{-1}\sum_{t=1}^{T} \Delta L_{t+1}\right)\hat{\Omega}^{-1}\left(T^{-1}\sum_{t=1}^{T} \Delta L_{t+1}\right), \tag{18}$$

where $\Delta L_{t+1} = (\bar{R}_{m,t} - R_{m,t+1})^2 - (\widehat{R}_{m,t+1|t} - R_{m,t+1})^2$ and $\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(\Delta L_{t+1} - \overline{\Delta L})^2$. Asymptotically, this Wald statistic follows a Chi-square distribution with degrees of freedom equal to the difference in the number of predictors.

Although this approach is commonly used, its performance may depend on which estimation period k the researcher chooses. This choice is sensitive for out-of-sample predictions since the R^2s of the in-sample predictive regressions are low most of the time. A low R^2 can be problematic since the standard errors of the regression are negatively related to the R^2.[11] To forecast more accurately, a longer sample (k) is needed for the in-sample estimation. However, the sample horizon cannot be too long if the predictive relationship is thought to change rapidly over time.

The new approach deviates in one critical dimension. The OOS forecast of month $T+1$ returns is formed by using the negative of the contemporaneous variance beta from month T in place of the predictive beta estimated over the past k periods. For the time being, I set the intercept of the predictive relation equal to zero. The one-step-ahead predicted value of market excess returns is then

$$\widehat{R}_{m,T+1|T} = -\hat{\beta}_{v,T} VRP_T. \tag{19}$$

I term this the "contemporaneous beta" approach because it directly uses the contemporaneous betas ($\hat{\beta}_{v,T}$) estimated from regressions of returns and variance innovations. Recall that the market risk premium consists of two parts. The premium that comes from the variance shock is equivalent to the product of the contemporaneous beta and the VRP. The premium that originates from the orthogonal shock may also be related to the VRP, but whether or how much it is related to the VRP is unknown. If either orthogonal risk is unpriced or its price is unrelated to the VRP, I expect the market risk premium to be related to the VRP with a slope that equals the size of the contemporaneous beta.

Since the betas are estimated from a single month of data, the forecast naturally uses the most up-to-date information on the market. There are several other benefits of doing this. Above all, the R^2s of the contemporaneous regressions, estimated using daily data, are typically higher than those of the historical predictive regressions, estimated using monthly observations. A higher R^2 allows using a shorter time period. Therefore, even when there is a change in the economic conditions or a structural break, the influence may be much smaller. Also, since the contemporaneous correlation is related to the strength of the predictive relationship, under the new approach, one can choose to use the information embedded in the VRP selectively, only when the premium is likely to be more informative about the market risk premium.

In the contemporaneous beta approach, the product of the negative variance risk exposure and the VRP predicts excess market returns with a zero intercept. This relation is based on the assumption that the orthogonal component is either unpriced or is too noisy to determine in the short-run. If the orthogonal component is priced, forming OOS forecasts based only on the VRP could result in poor forecasts of the equity risk premium. The orthogonal premium may comove with the VRP since both of them are prices of risk that depend on aggregate risk aversion. When the two prices are highly correlated over time, the contemporaneous beta may provide a biased estimate of the predictive slope.

Therefore, I evaluate the possibility that the orthogonal premium is explained by other well-known predictors of market returns, such as the dividend yield, that is related to a more persistent component of the risk premium. Therefore, I consider a third approach, a combination of the contemporaneous beta and traditional approaches. I use predictors of market returns that are known to perform well. What I call the "hybrid approach" is designed such that the orthogonal premium is allowed to be a linear function of common predictors. To do so, after estimating the contemporaneous betas from a first-stage regression in each month, I run a second regression,

$$R_{m,t+1} = -\hat{\beta}_{v,t} VRP_t + \delta_0 + \delta_1 \sqrt{1-\hat{\rho}_t^2} X_t + \eta_{t+1}, \tag{20}$$

to find estimates of $\hat{\delta}_0$ and $\hat{\delta}_1$ on a rolling basis. Here, X_t can be any predictor of market returns, including the VRP. Under this approach, the OOS forecast at time T is then,

$$\hat{R}_{m,T+1|T} = -\hat{\beta}_{v,T} VRP_T + \hat{\delta}_0 + \hat{\delta}_1 \sqrt{1-\hat{\rho}_T^2} X_T. \tag{21}$$

The forecasts of the hybrid approach are intended to incorporate the risk premium associated with the orthogonal risk component in addition to the variance risk component. Since the latter is well-captured by the product of the exposure and the price of variance risk, the role of this additional predictor (X_T) is limited to explaining the orthogonal premium. The term in the square-root ensures that the predictors are weighted conditionally depending on the relative size of the orthogonal component to total market risk.

It is possible that the orthogonal premium is not well captured by any of the predictors considered. There is also the possibility that the premium on orthogonal risk does not vary much over time or even remain constant. Hence, I consider a restricted case of the above approach, where δ_1 term is dropped. This particular case will be referred to as the contemporaneous beta approach with an intercept.

Each of the OOS-R^2 values is computed by comparing the performance of the one-step-ahead prediction of the model to the historical average. I use the rolling window of the past ten years of data to compute the historical mean as a benchmark. As shown in the robustness section, the results do not change much if the estimation period for the historical mean is restricted to a shorter period, or when some of the pre-1990 sample is included.

Table 2 summarizes the OOS-R^2s and the Wald statistics, along with p-values, for the different methods discussed. I mainly consider the sample that starts from 1993 because the traditional way of return forecasting requires that the estimates from in-sample regressions be used to form an OOS forecast. To show that the outperformance of the new methodology is not driven by a short sample used for in-sample regressions, in the robustness section, I also re-do the analysis using a sample that begins in 1998.

While Goyal and Welch (2008) show that returns are unpredictable OOS for most predictors proposed, Campbell and Thompson (2008) conjecture that the results may look different if we constrain the coefficients and slope of the

[11] The standard deviation of the prediction error, the mean-squared prediction error (MSPE) in a simple linear regression is,

$$MSPE = \sigma \sqrt{\frac{\sum_t (R_{m,t} - \hat{R}_{m,t})^2}{k}},$$

where $\sigma^2 = \mathrm{Var}(\epsilon_{p,t})$. The above is equivalent to $\sigma\sqrt{(1-R^2)\widehat{\mathrm{Var}}(R_{m,t})}$.

Table 2
Out-of-sample performance evaluation (1993–2016).

This table summarizes the out-of-sample predictive performance for one-month market returns over 1993–2016, using the VRP, estimated in one of several ways as described in Table 1, as a predictor.

The traditional approach uses the past ten years of data to run a first-stage predictive regression of monthly market returns on the lagged VRP. A one-step-ahead forecast is then formed using these coefficients and compared with the actual realized value. The contemporaneous beta approach sets the forecast equal to the negative of the contemporaneous beta multiplied by the VRP. The risk-free rate is added to make it comparable to actual returns. The forecast of the contemporaneous beta approach including intercept is obtained by estimating δ_0 from the regression

$$R_{m,t+1} = -\hat{\beta}_{v,t} VRP_t + \delta_0 + \eta_{t+1} \tag{32}$$

and forming a forecast based on the estimates. The out-of-sample R^2s and Wald statistics along with the p-values of the statistics are provided for each of these methods. The constrained forecasts are formed by setting any negative return forecasts to zero.

		VRP measures					
		VRP_N		$VRP_{\overline{P}}$		VRP_{P^E}	
		Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained	Constrained
Panel A: The traditional approach							
	OOS-R^2	0.010	−0.007	0.002	−0.008	0.052	0.032
	Wald	0.064	0.033	0.004	0.060	1.501	0.875
	p-value	(0.800)	(0.857)	(0.952)	(0.807)	(0.221)	(0.350)
Panel B: The contemporaneous beta approach							
B-1: No intercept							
1-month	OOS-R^2	0.065	0.061	0.079	0.076	0.084	0.079
WLS	Wald	5.027	6.836	7.956	10.838	5.996	6.984
	p-value	(0.025)	(0.009)	(0.005)	(0.001)	(0.014)	(0.008)
1-month	OOS-R^2	0.054	0.051	0.068	0.066	0.085	0.069
OLS	Wald	3.686	5.325	8.056	8.964	4.544	5.665
	p-value	(0.055)	(0.021)	(0.005)	(0.003)	(0.033)	(0.017)
3-month	OOS-R^2	0.064	0.066	0.049	0.052	0.064	0.064
WLS	Wald	11.460	12.232	3.554	3.818	4.130	4.122
	p-value	(0.001)	(0.000)	(0.059)	(0.051)	(0.042)	(0.042)
3-month	OOS-R^2	0.053	0.060	0.041	0.047	0.054	0.059
OLS	Wald	10.096	11.462	2.735	3.552	3.339	3.862
	p-value	(0.001)	(0.001)	(0.098)	(0.059)	(0.068)	(0.049)
B-2: Including intercept							
1-month	OOS-R^2	0.059	0.053	0.074	0.068	0.080	0.073
WLS	Wald	4.078	5.317	3.668	8.991	5.096	6.018
	p-value	(0.043)	(0.021)	(0.055)	(0.003)	(0.024)	(0.014)
1-month	OOS-R^2	0.047	0.045	0.062	0.061	0.065	0.064
OLS	Wald	4.920	4.084	6.485	7.543	5.096	4.899
	p-value	(0.027)	(0.043)	(0.011)	(0.006)	(0.024)	(0.027)

predictive regressions. They show that the forecast performance sometimes increases when a constraint is imposed. Motivated by their study, I impose a positivity constraint on the return forecasts for both the traditional approach of return forecasting and the new proposed methodology. The forecast without the positivity constraints is denoted by “Unconstrained” in the table, while the one with the constraint is denoted by “Constrained” in the table.

When applying the traditional approach, the VRP predicts market returns with a slightly positive OOS-R^2. The nonparametric VRP has a slightly positive OOS-R^2 of 1.0%. Consistent with Bekaert and Hoerova (2014), the predictive performance of the VRP largely depends on how the VRP is measured. The OOS-R^2 is 0.2% when the VRP is measured using the monthly average but increases to 5.2% when end-of-month values are used. The difference in the performance is largely due to two extreme observations during the financial crisis. This is because in September 2008, the VRP is estimated to be negative, which is followed by a large negative shock (−17.2%) in the index. Then, in October of 2008, there is a positive spike in the VRP, which is followed by a negative shock (−7.8%) in the index. Although not reported in the table, the gap among the three OOS-R^2s decreases to 2.0% (the R^2s in between 1.6% and 3.6%), when these two observations are removed. The influence of the first outlier is confirmed when the positivity constraint is imposed in the forecast. The OOS-R^2s decrease across all three measures with the biggest impact on the parametric VRP measured at the end of the month. Overall, despite sometimes having positive OOS-R^2s, none of the predictions of the traditional approach is statistically significant, even at the 10% level. The Wald statistics are statistically insignificant, and I conclude that there is no predictability in monthly market returns.

However, the numbers look much different when combining the contemporaneous beta with the VRP. The results when the orthogonal premium is set to zero (“no intercept”) are provided in Panel B-1. The OOS-R^2 for the parametric VRP is 7.9%–8.4% if multiplied by the WLS beta and 6.8%–8.5% if combined with the OLS beta, which is much higher than the numbers of the traditional approach. For the nonparametric VRP, the OOS-R^2 is 6.5% if the WLS beta is used and 5.4% when the OLS beta is used. For any given combination of the VRP and the beta estimates, the forecasts are at least 3% higher than the counterpart of the traditional approach. Most of all, the Wald statistics are now mostly significant. All the combinations considered are statistically significant at the 10% level. At the 5% level, five out of six specifications are statistically significant. The prediction error is on average small, and it varies little over time. These results suggest that using the new proposed approach, predicting one-month market returns in a statistically significant manner is possible, even out of sample. Using the constrained estimate does not change the result. The OOS-R^2s decrease by a small amount, but the Wald statistics increase even further, leading to statistical significance at 1% across four out of six combinations. The OOS-R^2s are still much higher than those of the traditional approach.

Using the three-month moving average beta sometimes increases but also decreases the forecast accuracy. The next part of the panel shows this. Generally, the nonparametric VRP performs better when combined with the three-month betas, while the parametric measures perform better with one-month betas. The Wald statistics are all significant regardless of the sample length used for the beta estimation.

When the orthogonal premium is assumed to be a nonzero constant ("including intercept"), the OOS-R^2 slightly decreases but generally stays at a similar level. The Wald statistics apparently decrease. This is because while assuming a constant for the orthogonal premium increases the fit for it, additional estimation error is added for the OOS forecast. Therefore, the fact that the OOS forecast does not decrease much may not necessarily suggest that the orthogonal risk premium is trivial, and the VRP only captures a fraction of the total market risk premium.

To better understand when the new approach performs especially better over the traditional approach, I develop a measure that computes the cumulative improvements in the loss function over the benchmark. I define the Cumulative Outperformance of the Forecast (COF) as:

$$COF_T = \sum_{t=1}^{T} \Delta L_t, \tag{22}$$

where L_t is the square loss function given in Eq. (18). Fig. 4 plots the COF of the constrained forecast, and Fig. 5 that of the unconstrained forecast. The dotted line shows the performance of the traditional approach, and the solid line shows the estimates using the one-month WLS beta. The time-series plot has an upward trend when the forecast performs better than the benchmark. A higher slope in a short time indicates that the predictions were more accurate during that particular time, and a higher level means that the forecast methodology performs well.

There are two things to note from these figures. First, these two figures suggest that the new approach strictly dominates the forecast of the traditional approach. The difference between these two prediction methodologies is especially pronounced for the post-1998 period. This is partly expected from Fig. 2 because these are also periods when the time-variation in the contemporaneous betas is more clearly observed. Second, the figures suggest that, unlike the traditional approach, the outperformance of combining the contemporaneous beta approach is not driven by a single observation or a very short period.

4.2. Time-varying out-of-sample predictability

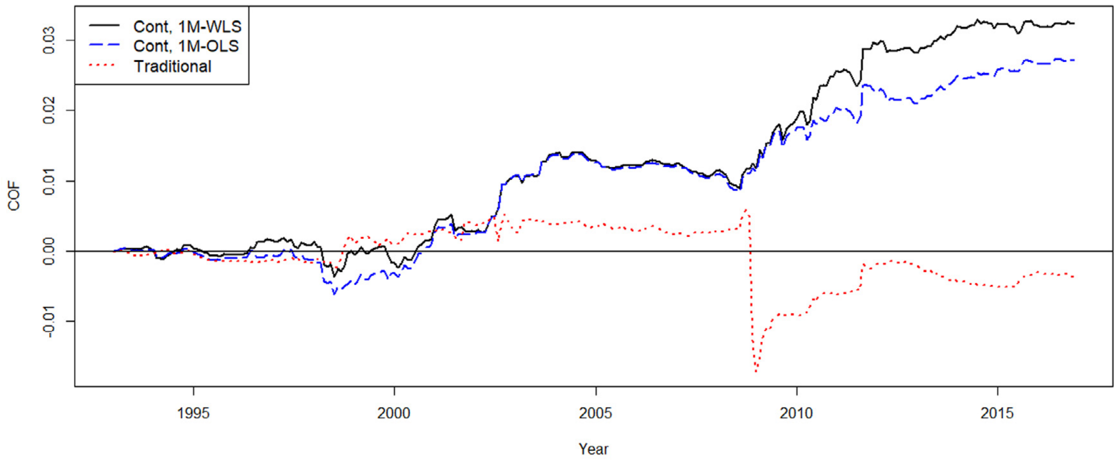
I also study the connection between contemporaneous correlations and predictive R^2s. I do so by dividing the full sample into different non-overlapping subsamples. Each of the 288 months in the full sample period of 1993–2016 is classified into one of three groups according to the monthly series of the contemporaneous correlations between market returns and variance innovations. When the correlation during a particular month is more negative than the first tercile of the historical distribution of past values, the month is classified as a "high" month. When the correlation is more positive than the second tercile, it is classified as a "lower" month. Otherwise, it is classified as a "medium" month. Therefore, the classifications are made without any look-ahead bias. Then, the OOS-R^2s are computed separately for each of these groups. For the remainder of this section, I focus on the unconstrained forecasts.

Table 3 summarizes the OOS-R^2s for each of the subsamples. For the contemporaneous beta approach, the same estimation interval is used for both correlations and betas. That is, when one-month betas are used for predictions, the sample classifications are also based on one-month correlations. I provide only the OOS-R^2s and not the Wald statistics due to shorter sample periods, so these results should be viewed somewhat informally.

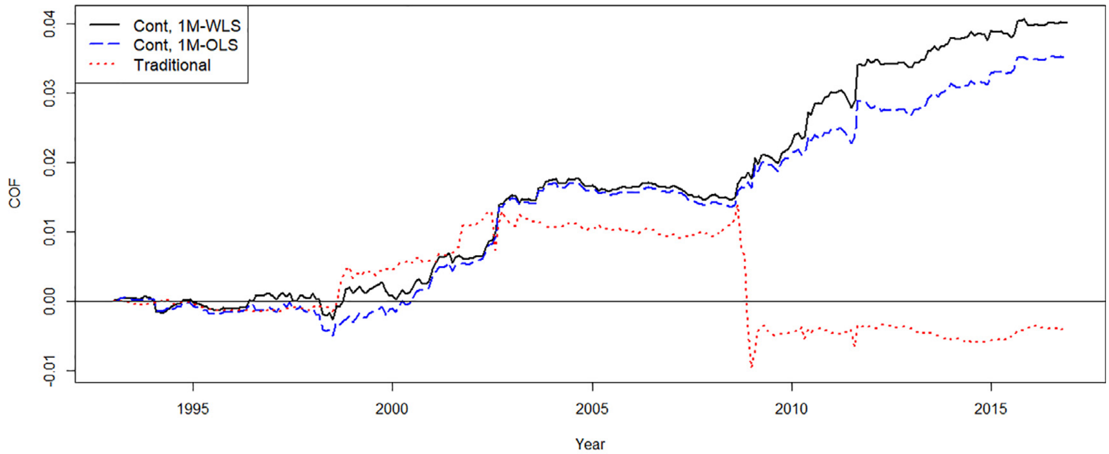
For the traditional approach, there is a substantial difference in OOS predictability. The OOS-R^2 between high and low periods is 16.4%–19.3% if the classifications are based on one-month correlations. For the classifications using three-month estimates, the difference is 14.4%–29.8%. This difference is much smaller for the contemporaneous beta approach. The differences between the two periods are 3.2%–13.6% if the sample is classified based on one-month correlations and 0.6%–8.8% for three-month correlations.

The smaller dispersion between high and low periods is consistent with the model. During low correlation times, the traditional method of running rolling predictive regressions effectively overstates the role of the VRP by assuming a constant predictive coefficient. However, this is not the case for the contemporaneous beta approach. The new approach already embeds information about the relation between returns and market variance in the beta, so that it does not rely excessively on the VRP during low correlation periods.

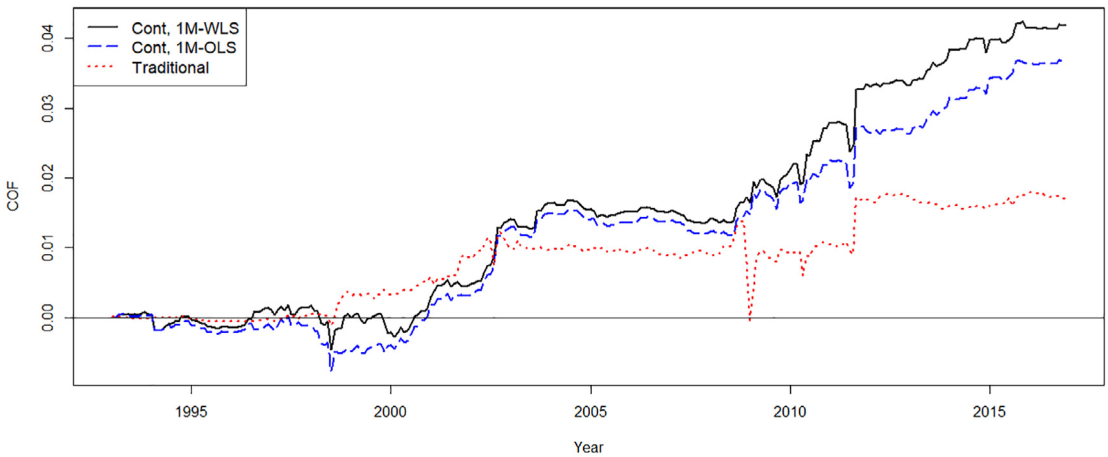
Thus, when market prices and variance move closely together, the VRP is a very powerful predictor of short-horizon market returns. On the other hand, when they move independently, it is hard to predict market returns using the VRP, since the market portfolio is less exposed to variance risk.



(a) Variance risk premium (nonparametric)



(b) Variance risk premium (parametric, averaged)

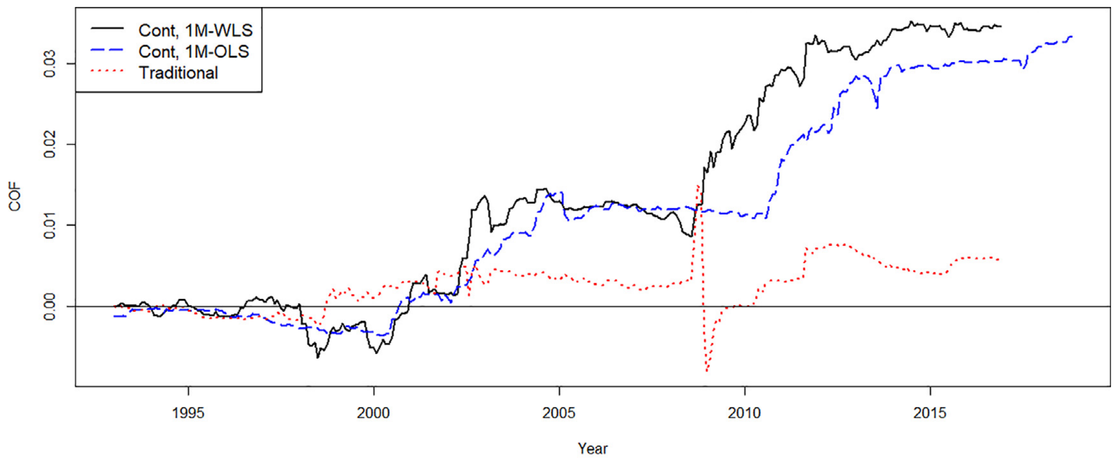


(c) Variance risk premium (parametric, end of month)

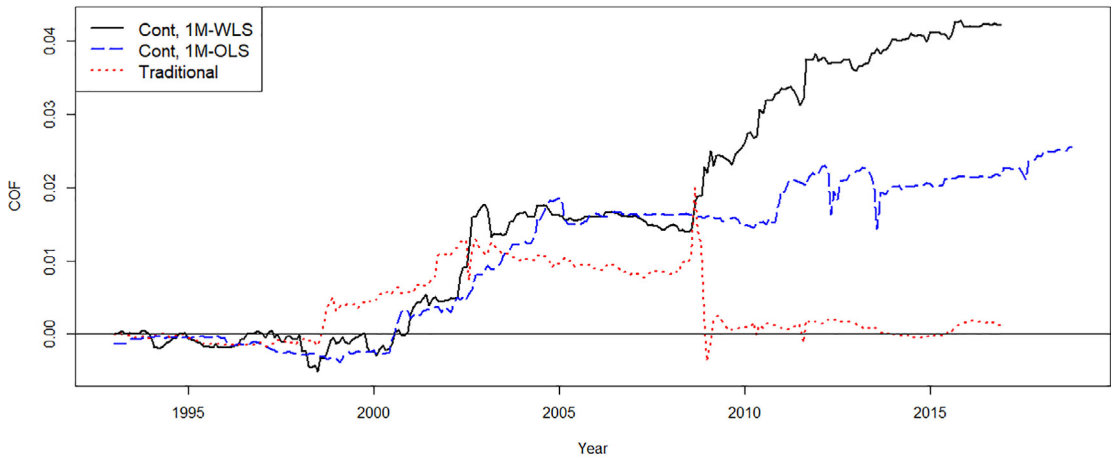
Fig. 4. Comparing the out-of-sample performance (constrained). This figure compares the out-of-sample performance between the traditional approach of return forecasting and the contemporaneous beta approach, when the VRP is used as a monthly return predictor. The figure plots the time series of the Cumulative Outperformance of the Forecast (COF) defined as

$$\mathrm{COF}_T = \sum_{t=1}^{T}[(\bar{R}_{m,t} - R_{m,t+1})^2 - (\max(\widehat{R}_{m,t+1|t}, 0) - R_{m,t+1})^2]. \tag{29}$$

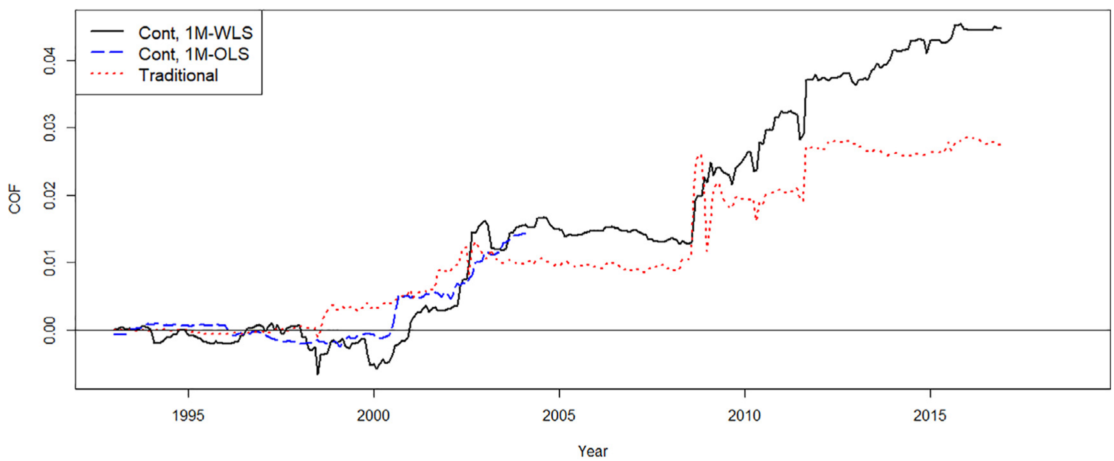
The COF of the traditional approach is shown in dotted lines, the one for the contemporaneous beta approach with one-month WLS betas in solid lines, and the one with one-month OLS betas in dashed lines.



(a) Variance risk premium (nonparametric)



(b) Variance risk premium (parametric, averaged)



(c) Variance risk premium (parametric, end of month)

Fig. 5. Comparing the out-of-sample performance (unconstrained). This figure compares the out-of-sample performance between the traditional approach of return forecasting and the contemporaneous beta approach when the VRP is used as a monthly return predictor. The figure plots the time series of the Cumulative Outperformance of the Forecast (COF) defined as

$$\mathrm{COF}_T = \sum_{t=1}^{T} (\bar{R}_{m,t} - R_{m,t+1})^2 - (\widehat{R}_{m,t+1|t} - R_{m,t+1})^2. \tag{30}$$

The COF of the traditional approach is shown in dotted lines, the one for the contemporaneous beta approach with one-month WLS betas in solid lines, and the one with one-month OLS betas in dashed lines.

Table 3
Conditional out-of-sample predictions.
This table compares the out-of-sample performance of one-month market return predictions using the VRP as a predictor across different samples. Each month in the sample is classified as a high, medium, or a low month depending on whether the contemporaneous correlation is greater or less than the historical terciles evaluated using data available up to that point. The contemporaneous correlation is defined as the correlation between daily market returns ($R_{m,\tau}$) and variance innovations ($RV_\tau - E_{\tau-1}[RV_\tau]$). 'High' contains observations for which the estimates are highly negative. The classification is based either on one-month or on three-month correlations. The estimation period for the contemporaneous betas is matched to that of the correlations. Then, the out-of-sample R^2s are evaluated for each of these subsamples separately.

		OOS-R^2					
		VRP_N		$VRP_{\overline{P}}$		VRP_{P^E}	
		1-month	3-month	1-month	3-month	1-month	3-month
Panel A: The traditional approach							
	High	0.068	0.162	0.096	0.065	0.164	0.207
	Medium	0.049	0.069	−0.035	0.057	−0.050	0.060
	Low	−0.124	−0.135	−0.097	−0.079	0.000	−0.056
	High-low	0.193	0.298	0.193	0.144	0.164	0.263
Panel B: The contemporaneous beta approach							
B-1: No intercept							
WLS	High	0.082	0.131	0.122	0.065	0.161	0.128
	Medium	0.056	0.054	0.053	0.048	0.031	0.053
	Low	0.050	0.047	0.045	0.059	0.036	0.048
	High-low	0.032	0.085	0.077	0.006	0.125	0.080
OLS	High	0.079	0.126	0.120	0.064	0.156	0.123
	Medium	0.042	0.042	0.038	0.037	0.009	0.041
	Low	0.029	0.039	0.024	0.050	0.020	0.040
	High-low	0.050	0.088	0.096	0.014	0.136	0.083
B-2: Including intercept							
WLS	High	0.074	0.147	0.114	0.084	0.139	0.149
	Medium	0.057	0.033	0.055	0.024	0.050	0.031
	Low	0.040	0.035	0.035	0.050	0.024	0.038
	High-low	0.034	0.112	0.079	0.034	0.115	0.111
OLS	High	0.070	0.140	0.111	0.080	0.131	0.141
	Medium	0.041	0.021	0.037	0.014	0.032	0.019
	Low	0.021	0.025	0.016	0.039	0.004	0.028
	High-low	0.049	0.114	0.095	0.040	0.128	0.112

4.3. *Explaining the orthogonal premium*

Return predictors other than the VRP may also complement the VRP for two reasons. First, the predictive power of the VRP is strong for monthly and quarterly returns. However, for other predictors, that power is higher for predictions of longer horizon returns (Poterba and Summers, 1988; Fama and French, 1988b). Second, the predictive strength of many common predictors tends to decrease for the post-1993 period.[12] In contrast, the VRP has been demonstrated to be a strong predictor of market returns in the post-1990 period. Given these differences, I hypothesize that the other predictors may help explain the risk premium that arises from orthogonal risk.

I select several predictors that are well-known to predict market returns. These include: dividend yield (D/Y) (Campbell and Shiller, 1988; Fama and French, 1988a), the term premium (TERM) (Campbell, 1987), the default premium (DEF) (Keim and Stambaugh, 1986), the short rate (Campbell, 1987), short interest (Rapach et al., 2016), *cay* (Lettau and Ludvigsen, 2001), and new orders-to-shipment (NO/S) (Jones and Tuzel, 2013). All of these are known to perform well in predicting market returns over a long historical sample. For completeness, I also consider the left jump variation (LJV) from Bollerslev et al. (2015) as a potential candidate. I also let the VRP itself explain variation in the orthogonal premium. Including the VRP is important because if the orthogonal premium and the VRP are related, it will alter how the VRP relates to expected market returns. In this case, the contemporaneous beta would be a biased estimator of the predictive slope.

Table 4 provides the OOS predictive performance of the hybrid approach, where the price of orthogonal risk is allowed to be time-varying. The left three columns evaluate the performance of the post-1993 sample. Only the results with the one-month OLS beta are summarized as OLS is easier to use, but the results are not much different even when using WLS instead. The first two rows repeat

[12] See, for example, Goetzmann and Jorion (1993), Ang and Bekaert (2007), and Goyal and Welch (2008).

Table 4
Analyzing the price of orthogonal risk.
This table provides the out-of-sample performance of the hybrid approach, in which additional predictors are used to complement the contemporaneous beta approach. Forecasts are formed from the regression

$$R_{m,t+1} = -\hat{\beta}_{v,t} VRP_t + \delta_0 + \delta_1 \sqrt{1-\hat{\rho}_t^2} X_t + \eta_{t+1}, \tag{33}$$

where X_t is some predictor studied in previous literature. The additional predictors are described in Section 4.3. The out-of-sample R^2s and the p-values are provided for each set of forecasts. Also, in the right four columns, the out-of-sample R^2s are provided for each of the subsample classifications. To save space, for the subsample analysis, only the results for the nonparametric VRP are provided.

Additional		1993–2016			Subsamples			
variables (X_t)		VRP_N	$VRP_{\overline{P}}$	VRP_{P^E}	High	Medium	Low	H-L
Intercept only	OOS-R^2	0.047	0.062	0.065	0.070	0.041	0.021	0.049
	p-value	(0.027)	(0.011)	(0.024)				
D/Y	OOS-R^2	0.016	0.042	0.033	0.017	0.019	0.013	0.004
	p-value	(0.464)	(0.167)	(0.235)				
TERM	OOS-R^2	0.021	0.053	0.028	0.058	0.019	−0.021	0.079
	p-value	(0.260)	(0.061)	(0.234)				
DEF	OOS-R^2	−0.015	0.018	0.002	0.000	0.045	−0.079	0.079
	p-value	(0.639)	(0.614)	(0.962)				
Short rate	OOS-R^2	−0.010	0.008	-0.310	0.007	0.031	−0.063	0.069
	p-value	(0.731)	(0.839)	(1.000)				
Short interest	OOS-R^2	0.020	0.052	0.023	0.048	0.028	−0.020	0.068
	p-value	(0.361)	(0.096)	(0.397)				
cay	OOS-R^2	0.025	0.054	0.038	0.089	−0.067	0.023	0.066
	p-value	(0.359)	(0.125)	(0.240)				
NO/S	OOS-R^2	0.015	0.048	0.025	0.036	0.027	−0.019	0.055
	p-value	(0.434)	(0.106)	(0.300)				
LJV	OOS-R^2	−0.104	−0.189	−0.103	−0.013	−0.108	−0.212	0.199
	p-value	(0.092)	(0.276)	(0.205)				
VRP	OOS-R^2	−0.001	−0.070	−0.074	0.069	0.023	−0.100	0.169
	p-value	(0.995)	(0.312)	(0.150)				

the results from Table 2, where the orthogonal premium is assumed to be constant. These results will serve as a benchmark for determining whether the predictors help to explain the orthogonal premium. The right four columns summarize the OOS-R^2s for the high, medium, and low subsamples. Finally, the last column is the difference between the high and low periods. Only the results for the nonparametric VRP are provided, but the results are similar for other measures.

If evaluated over the entire sample period, none of the other eight predictors considered improves the OOS-R^2s over the benchmark. If the subsamples are analyzed separately, the default premium improves the predictability during medium periods, and *cay* improves the forecast during high periods. During low periods, *cay* improves the predictability. However, the magnitude of the overall improvement is small, and they may not be performing well in capturing the short-term risk premium, such as at the monthly level. Finally, when the VRP is used additionally to explain the variation in the orthogonal premium, the R^2s do not improve for the entire sample; the Wald statistics are lower, and the OOS-R^2s are even smaller. The last result indicates that there is no strong evidence that the orthogonal premium is related to the VRP.

4.4. Evaluating economic significance– a trading strategy

I also evaluate whether the closeness between the two betas can be used to form a trading strategy. Following Goyal and Welch (2008), I use the one-step-ahead OOS forecasts to calculate optimal weight on the stock market as

$$w_T = \frac{\hat{R}_{m,T+1|T}}{\gamma \hat{\sigma}_T^2}, \tag{23}$$

where $\gamma = 3$ is assumed for the risk aversion coefficient and the monthly square of VIX is used as a proxy for $\hat{\sigma}^2$. The remaining proportion $1 - w_T$ is invested in the risk-free asset. The weight in the market is capped at 200%. The certainty equivalent (CE) of the return is computed as

$$CE = \overline{R_p} - \frac{\gamma}{2}\widehat{\text{Var}}(R_p), \tag{24}$$

where $\overline{R}_p$ and $\widehat{\text{Var}}(R_p)$ are the sample mean and variance of the portfolios, respectively. I compare the Sharpe ratios (SRs) and the CEs of various forecasts with the baseline, in which the historical mean is assumed to be the best predictor of future returns.

The previous tables on the predictive performance show that predictions can be made more accurately when the absolute correlation between returns and variance is high. A concern is that the weights might rely too much on the VRP-based forecasts during periods when returns and variance innovations are unrelated. Therefore, I also consider an alternative strategy, in which a fraction of the allocation of stocks depends on the model-based predicted returns and the rest on the historical average of past returns. The weight invested in the risky asset becomes

$$w'_T = \frac{\hat{R}_{m,T+1|T}}{\gamma \hat{\sigma}_T^2}\sqrt{\hat{\rho}_T^2} + \frac{\overline{R}_{m,T}}{\gamma \hat{\sigma}_T^2}\sqrt{1-\hat{\rho}_T^2}, \tag{25}$$

Table 5
Evaluation of the economic significance of out-of-sample trading strategies.

This table summarizes the out-of-sample investment performance in terms of the annual Sharpe ratios (SR) and the certainty equivalents (CE), assuming a risk-averse investor with a risk-aversion coefficient of three. The weight of the market portfolio is determined by the out-of-sample forecasts constructed using a fixed weight, historical average, the traditional approach, and the contemporaneous beta approach, each of them capped at 0% and 200%. The rest $1 - w_T$ is invested in the risk-free asset. The 'unconditional' weights are formed in a way that does not rely on the contemporaneous correlations. The weights are

$$w_T = \frac{\widehat{R}_{m,T+1|T}}{\gamma \hat{\sigma}_T^2}. \tag{34}$$

The 'conditional' weights depend on the size of the estimated contemporaneous correlations ($\hat{\rho}$) and equal

$$w_T' = \frac{\widehat{R}_{m,T+1|T}}{\gamma \hat{\sigma}_T^2}\sqrt{\hat{\rho}_T^2} + \frac{\overline{R}_{m,T}}{\gamma \hat{\sigma}_T^2}\sqrt{1-\hat{\rho}_T^2}. \tag{35}$$

The benchmark is when the historical average is used as a forecast for the next month. Gains/Losses relative to the benchmark are reported.

	Unconditional weighting				Conditional weighting			
	SR	CE	Δ SR	Δ CE	SR	CE	Δ SR	Δ CE
Fixed weight	0.527	0.046						
Average (benchmark)	0.632	0.040						
The traditional approach								
VRP_N	0.524	0.033	−0.108	−0.007	0.673	0.048	+0.042	+0.009
$VRP_{\overline{P}}$	0.667	0.044	+0.035	+0.004	0.739	0.054	+0.107	+0.014
VRP_{P^E}	0.672	0.053	+0.040	+0.013	0.741	0.059	+0.109	+0.019
The contemporaneous beta approach with no intercept								
VRP_N	0.760	0.078	+0.129	+0.039	0.729	0.071	+0.097	+0.031
$VRP_{\overline{P}}$	0.922	0.098	+0.290	+0.058	0.836	0.084	+0.204	+0.044
VRP_{P^E}	0.782	0.090	+0.151	+0.050	0.749	0.078	+0.117	+0.039
The contemporaneous beta approach including intercept								
VRP_N	0.753	0.081	+0.121	+0.041	0.715	0.070	+0.083	+0.030
$VRP_{\overline{P}}$	0.902	0.098	+0.270	+0.059	0.817	0.081	+0.185	+0.042
VRP_{P^E}	0.812	0.097	+0.180	+0.057	0.750	0.079	+0.119	+0.039

where $\hat{\rho}_T$ is the estimated contemporaneous correlation between index returns and variance innovations of month T. To distinguish this strategy formed on the conditional value of the correlation from the basic trading strategy, I call this the conditional trading strategy and the first one as the unconditional trading strategy.

Table 5 summarizes the resulting gains/losses in the annualized SRs and CEs. All numbers are based on one-month return forecasts, and the statistics are annualized. The left two columns compare the SR and CE of the unconditional trading strategy. The first row summarizes the values of holding a fixed unit weight on the market portfolio. The next row is the performance of the portfolio, where the past average and VIX are used to form the weights on stocks and bonds. This trading strategy serves as a benchmark. Both the SR and CE are slightly higher than the fixed unit weight as it systematically de-weights the risky portfolio when market variance is high. The SRs and CEs are compared to the benchmark, and the relative gains or losses are reported in subsequent rows.

For the unconditional trading strategy, the gains in SR and CE are extremely small when predicting returns as in the traditional manner, but increase substantially when using the contemporaneous beta approach. Notably, compared to the benchmark, the performance of the traditional approach even decreases for one of the three VRP measures. Although there are slight gains in SRs and CEs using the other two measures, they are very small in magnitude. The gain in annual SR is less than 0.04 and is 1.3% in CE. On the other hand, the trading strategies based on the contemporaneous beta approach show a substantial improvement. The SRs increase by 0.129–0.151, and the CEs increase by 3.9%–5.8%. The gains in assuming a constant orthogonal premium are marginal compared to assuming no intercept. Some of the SRs increase, while others decrease.

The last four columns summarize the gains and losses in SRs and CEs for the conditional trading strategies. When using the traditional approach, the gains in SRs and CEs are slightly larger than those of the unconditional trading strategies. However, the performances are slightly worse for the conditional strategy compared to the unconditional strategy when incorporating the contemporaneous beta. As discussed, this is presumably because the contemporaneous beta already embeds information about the contemporaneous correlations. In any case, the performance of the contemporaneous beta approach still dominates that of the traditional approach. These results are also similar when including an intercept.

In conclusion, these results indicate that predictions under the traditional approach could be highly misleading during periods when returns and variance innovations are unrelated. During these times, investors appear to perceive variance risk as unrelated to market risk. The VRP, therefore, provides little information about the market risk

Table 6
Predictive regressions of monthly market returns (1990–2016).

This table summarizes the results of one-month in-sample predictive regressions of market returns ($R_{m,t+1}$) using the VRP as a return predictor. Columns 1, 4, and 7 show the results of the regression:

$$R_{m,t+1} = \beta_0 + \beta_p VRP_t + \epsilon_{t+1}. \tag{36}$$

The other columns summarize the results of the interactive predictive regression:

$$R_{m,t+1} = \gamma_0 + \gamma_p VRP_t + \gamma_I \hat{\beta}_{v,t} \times VRP_t + \epsilon_{t+1}, \tag{37}$$

where $\hat{\beta}_{v,t}$ is the contemporaneous beta that is either estimated from ordinary least squares (OLS) or weighted least squares (WLS) of one-month regression. Panel A summarizes the results when the predictive regressions are estimated using OLS, and Panel B shows the WLS results. The *t*-statistics estimated using heteroscedasticity-consistent standard errors are shown in parenthesis.

Panel A: Prediction using ordinary least squares

	VRP_N			$VRP_{\overline{P}}$			VRP_{P^E}		
$\hat{\beta}_{v,t}$		OLS	WLS		OLS	WLS		OLS	WLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VRP_t	4.485*	4.030*	3.931*	3.333*	2.542	2.396*	5.497***	4.210**	3.874**
	(1.85)	(1.82)	(1.80)	(1.65)	(1.31)	(1.84)	(3.15)	(2.18)	(1.96)
$VRP_t \times \hat{\beta}_{v,t}$		-0.751***	-0.843***		-0.883***	-0.973***		-0.579**	-0.660**
		(2.67)	(3.02)		(3.15)	(3.50)		(2.29)	(2.56)
Adj-R^2	0.028	0.054	0.062	0.017	0.051	0.060	0.056	0.073	0.079

Panel B: Prediction using weighted least squares

	VRP_N			$VRP_{\overline{P}}$			VRP_{P^E}		
$\hat{\beta}_{v,t}$		OLS	WLS		OLS	WLS		OLS	WLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VRP_t	3.697*	3.276*	3.096*	3.279**	2.556	2.376	4.922***	0.380**	3.472**
	(1.91)	(1.76)	(1.68)	(2.23)	(1.51)	(1.41)	(3.13)	(2.26)	(2.01)
$VRP_t \times \hat{\beta}_{v,t}$		-0.052**	-0.626**		-0.663***	-0.740***		-0.497**	-0.579**
		(2.31)	(2.64)		(2.75)	(3.06)		(2.23)	(2.55)
Adj-R^2	0.014	0.027	0.033	0.012	0.038	0.038	0.034	0.046	0.050

*** denotes significance at 1%, ** at 5%, and * at 10% level.

premium. On the other hand, when the correlation is highly negative, the VRP and the market risk premium are also highly related because market and variance risk are closely related. Moreover, they are connected in a particular way, so that the market's exposure to variance risk can replace the predictive beta. The contemporaneous beta approach predicts the one-month market also in an economically significant manner.

5. In-sample predictions

In this section, I confirm that the key results also hold in sample. I first summarize the results of the classical predictive regressions, replicating that of Bollerslev et al. (2009). Next, I examine properties of the time-varying predictive beta and whether the predictive beta can be inferred from the past contemporaneous relation between returns and variance innovations. Then, I show that the in-sample predictive beta is approximately proportional to the contemporaneous beta. Finally, I investigate the performance of the predictive regressions over time and demonstrate that their accuracy is related to the correlation between market returns and variance innovations.

Bollerslev et al. (2009) run a classical predictive regression of market excess returns on the VRP and find a positive and statistically significant predictive slope for horizons of less than six months. This paper suggests that this predictive beta may change over time. They must be higher when the market portfolio loads more on variance risk, and lower when the market does not load on variance risk. This hypothesis can be directly tested by running the regression of

$$R_{m,t+1} = \gamma_0 + \gamma_p VRP_t + \gamma_I VRP_t \times \hat{\beta}_{v,t} + \epsilon_{p,t+1}, \tag{26}$$

where $R_{m,t+1}$ is the one-month predictive market return. A negative and statistically significant interactive coefficient γ_I means that the predictive betas and the contemporaneous betas are negatively related, as hypothesized.

Johnson (2017) argues that for many predictive regressions, the predictive powers become insignificant when using generalized least squares instead. For the VRP, he shows that the predictive slope is insignificant with WLS if tested on the original sample of Bollerslev et al. (2009), 1990–2007. He claims that the return predictability if the VRP is driven by several extreme observations with high variance. To check for the possibility that using WLS may alter the results, in addition to OLS, I also consider WLS.

Table 6 summarizes the regression coefficients, *t*-statistics, and the adjusted-R^2s of the simple predictive regression as well as the interactive regressions. Panel A

Table 7
Conditional in-sample predictive performance.
This table summarizes the in-sample R^2s and t-statistics of monthly predictive regressions of one-month market returns ($R_{m,t+1}$) on the VRP for split samples. The regression is:

$$R_{m,t+1} = \beta_{0p} + \beta_p VRP_t + \epsilon_{t+1}. \tag{38}$$

Each month in the sample (1990–2016) is classified as a high, medium or a low month depending on whether the one-month contemporaneous correlation is greater or less than the historical terciles evaluated using data available up to that point. The contemporaneous correlation is defined as the correlation between daily market returns ($R_{m,\tau}$) and variance innovations ($RV_\tau - E_{\tau-1}[RV_\tau]$). 'High' contains samples in which the correlations are highly negative. The t-statistics, reported in parentheses, are adjusted for heteroscedasticity.

		Classification			
		High	Medium	Low	High-low
	Number of months	113	103	108	
VRP_N	In-sample R^2	0.117***	0.047**	0.004	0.113
		(3.83)	(2.26)	(0.69)	
	Predictive beta (β_p)	11.441	5.054	1.427	
$VRP_{\bar{P}}$	In-sample R^2	0.131***	−0.003	0.007	0.124
		(4.11)	(−0.12)	(0.78)	
	Predictive beta (β_p)	10.017	−0.282	1.580	
VRP_{pE}	In-sample R^2	0.179***	0.001	0.027	0.152
		(4.91)	(0.36)	(0.71)	
	Predictive beta (β_p)	8.743	3.936	0.026	

*** denotes significance at 1%, ** at 5%, and * at 10% level.

summarizes the results of OLS, and Panel B summarizes those of WLS. The simple predictive regression results are in Columns 1, 4, and 7 of the table, respectively. Among the three considered, the parametric VRP based on end-of-month observations performs best as a return predictor even in sample. This is similar to what is observed for OOS forecasts. First, the predictive slope of the OLS is between 3.33 and 5.50. Notably, these numbers are quite close to the average of contemporaneous variance betas reported in Table 1, −4.19, with the opposite sign. This is exactly what the model of this paper predicts. The WLS slope is also not much different in magnitude, although it is slightly smaller. Overall, the regression coefficients are statistically significant with similar t-statistics. The results are consistent with Johnson (2017), who shows that for the 1990–2015 sample, the level of significance of the VRP predictive regressions is unaffected by using WLS.

The other columns of the panel summarize the regression coefficients of the interactive regressions. Both the interactive regressions with OLS betas and WLS betas are presented. Across all measures of the contemporaneous beta estimate considered and for all VRP measures considered, the coefficient on the interactive variable is negative and statistically significant. On average, a single unit decrease in the one-month contemporaneous beta corresponds to an approximately 0.58–0.97 unit increase in the predictive betas. Although the interactive coefficients of the WLS regressions are slightly smaller, they are statistically significant with comparable t-statistics. Therefore, I conclude that the predictive and the contemporaneous betas are closely related to each other.

I also study the connection between contemporaneous correlations and predictive R^2s. I follow the OOS classification methodology considered in the previous section. The entire sample of 324 months is classified into three groups. High denotes months that has the most negative correlation between market returns and variance innovations. To avoid misclassification during the early years when there are not enough data for the benchmark distribution, I use the sample 1990–1994 to classify the first 48 months.

Next, I run a constant coefficient predictive regression for each of these subsamples separately. Table 7 reports the R^2s, coefficients, and t-statistics for each of the predictive regressions run separately for subsamples. The model suggests that the VRP should explain a greater fraction of the market risk premium when a large proportion of market risk is explained by variance risk. If this hypothesis is true, the R^2s must be high and statistically significant during high periods. On the contrary, during low periods, the R^2s of the predictive regressions should be insignificant. The empirical results support this hypothesis. The in-sample R^2s are much higher during high periods, 9.1%–14.0%, compared to low periods, which have R^2s of 0.4%–1.4%. Conclusively, this table suggests that there are times when the predictability of the VRP is stronger than other times, and we know, ex ante, when it should be stronger. Overall, the predictive power depends on the relationship between returns and variance innovations.

In short, the results show that the contemporaneous and predictive relations are linked in a very specific manner, such that the predictive beta depends on the contemporaneous beta. Moreover, the predictive performance, measured by R^2, increases as the correlation between market returns and variance innovations becomes more negative.

Table 8
Alternative VRP measures.
This table repeats the analyses of Tables 2 and 3 for three additional VRP measures. VRP_{BTZ} is from Bollerslev et al. (2009), and VRP_{BH} is from Bekaert and Hoerova (2014) estimated at the end of the month. Finally, VRP_{VXO} denotes the case in which both option-implied variance (VXO) and high-frequency realized variance are estimated using the S&P 100 Index.

Panel A: OOS performance of the traditional approach using alternative measures

	VRP_{BTZ}	VRP_{BH}	$VRP_{VXO,N}$		$VRP_{VXO,\bar{P}}$		VRP_{VXO,P^E}	
	1993–2016	1993–2016	1991–2016	1993–2016	1991–2016	1993–2016	1991–2016	1993–2016
OOS-R^2	0.037	0.024	−0.015	−0.009	−0.010	−0.007	0.011	0.016
Wald	2.176	0.367	0.233	0.082	0.233	0.082	0.186	0.339
p-value	(0.140)	(0.545)	(0.629)	(0.775)	(0.629)	(0.775)	(0.666)	(0.561)

Panel B: OOS performance of the contemporaneous beta approach

	Statistics	VRP_{BTZ}	VRP_{BH}	$VRP_{VXO,N}$		$VRP_{VXO,\bar{P}}$		VRP_{VXO,P^E}	
		1993–2016	1993–2016	1991–2016	1993–2016	1991–2016	1993–2016	1991–2016	1993–2016
One-month	OOS-R^2	0.050	0.073	0.068	0.070	0.083	0.086	0.084	0.087
WLS	Wald	2.245	3.877	5.993	5.723	9.145	8.781	6.906	6.666
	p-value	(0.134)	(0.049)	(0.014)	(0.017)	(0.002)	(0.003)	(0.009)	(0.010)
One-month	OOS-R^2	0.041	0.054	0.055	0.057	0.071	0.073	0.072	0.073
OLS	Wald	2.907	2.863	4.098	3.960	6.885	6.602	4.982	4.619
	p-value	(0.088)	(0.091)	(0.043)	(0.047)	(0.009)	(0.010)	(0.026)	(0.032)
Three-month	OOS-R^2	0.060	0.072	0.059	0.064	0.046	0.049	0.053	0.054
WLS	Wald	4.130	4.951	10.298	10.890	3.404	3.568	3.063	2.979
	p-value	(0.042)	(0.026)	(0.001)	(0.001)	(0.065)	(0.059)	(0.080)	(0.084)
Three-month	OOS-R^2	0.052	0.057	0.049	0.053	0.041	0.044	0.047	0.049
OLS	Wald	4.574	3.997	7.637	8.120	2.898	3.034	2.649	2.516
	p-value	(0.032)	(0.046)	(0.006)	(0.004)	(0.089)	(0.082)	(0.104)	(0.113)

Panel C: Conditional OOS performance (traditional approach, 1993–2016)

	OOS-R^2				
	VRP_{BTZ}	VRP_{BH}	$VRP_{VXO,N}$	$VRP_{VXO,\bar{P}}$	VRP_{VXO,P^E}
C-1: One-month correlations					
High	0.060	0.129	0.054	0.062	0.083
Medium	0.043	−0.057	0.045	0.011	−0.001
Low	0.002	−0.002	−0.124	−0.101	−0.049
High-low	0.058	0.131	0.178	0.163	0.131
C-2: Three-month correlations					
High	0.135	0.121	0.075	0.036	0.078
Medium	0.060	0.034	0.034	0.039	0.027
Low	−0.041	−0.040	−0.121	−0.089	−0.047
High-low	0.176	0.161	0.196	0.125	0.125

6. Robustness

6.1. Alternative measures of the variance risk premium

Bekaert and Hoerova (2014) argue that the return predictability of the VRP largely depends on how the VRP is measured. To understand how using a different variance forecast model affects the interactions of the contemporaneous beta and the VRP in explaining future market returns, I construct several other measures of VRP that have been used in previous research.

This paper is motivated by the predictive regression of Bollerslev et al. (2009), which reports a positive relation between the VRP and future market returns. The VRP in the article is constructed by taking the difference between the end-of-month value of implied variance and the monthly average of RV. As discussed, this timing mismatch could result in biases when market variance trends during a month. I consider the measure of Bollerslev et al. (2009) and denote this by VRP_{BTZ}.

I also consider the measure of Bekaert and Hoerova (2014). In this specification, the real-world expectation component of the VRP is measured using an alternative forecast model

$$\widetilde{RV}_t = \alpha_0 + \alpha_1 \widetilde{RV}_{t-1} + \alpha_2 VIX_{t-1}^2 + e_t, \tag{27}$$

where $\widetilde{RV}_t$ is the monthly sum of the daily RVs. This measure is commonly used (e.g., Gonzalez-Urteaga and Rubio, 2016; Chen et al., 2018) because both components are forward looking. However, this measure may still be biased to trending variance unless both components are measured at the same time. Therefore, I slightly modify their original measure and let VRP_{BH} be the difference between the end-of-month value of VIX and the RV forecast of the above model.

Finally, I construct a VRP measure using the S&P 100 Index. This alternative measure is useful because the option-implied variance on this index (VXO) is available from 1986. The VXO is matched with the S&P 100 Index realized variance, computed in the same manner as those of the S&P 500 Index. The VRP constructed in this manner, denoted by VRP_{VXO}, is also measured in three different ways. Due to the availability of the intraday trading data of the S&P 100 index, this measure is restricted to the sample period of 1988–2016. The variance betas and correlations are also computed using the realized variance of the S&P 100 Index, whenever one of these are used for prediction. Since this measure can be constructed for two additional years, it allows to rely on a slightly longer sample period.

Table 8 reports the key results of this paper using these alternative measures of the VRP. Overall, the results are similar to those of previous sections. Panel A shows the results of the traditional approach. This panel suggests that the OOS predictability of the VRP largely depends on how the VRP is measured. None of the Wald statistics are significant, suggesting that one-month market returns are non-predictable. Especially, adding the earlier sample of 1991–1993 makes the VRP a slightly worse predictor when using the traditional approach of return forecasting.

Panel B provides the OOS performance of the contemporaneous beta approach without an intercept. To save space, no intercept is assumed, but the results are similar assuming a constant orthogonal premium (i.e., similar R^2s with slightly lower Wald statistics). Across every measure considered, there is an increase in the OOS-R^2. For the BTZ measure, the OOS performance is not much different from the traditional approach when using a single-month beta but improves substantially when using the three-month beta. Part of the weak result may be because the BTZ measure contains a volatility trend that is correlated with the leverage effect. The result of the BH measure strengthens this explanation. Comparing the performance to the traditional approach, the OOS-R^2 increases by 3.0%–4.9%, and the Wald statistics all become statistically significant at the 10% level. However, although not reported in the table, the outperformance becomes much weaker if the BH measure is constructed to contain volatility trends. Finally, the VXO-based measure suggests monthly market returns are predictable even when the S&P 100 Index is used instead, or if two additional years of data is added.[13] All combinations of VRP and beta estimates considered can predict the monthly market returns significantly, except for one, when combining the end-of-month-based VRP with the three-month OLS beta.

Panel C summarizes the results of conditional OOS return forecasting. Only the results of the traditional approach for the post-1993 period are presented because the difference between high and low periods is expected to be strongest for the traditional approach. Regardless of the exact specification considered, returns are predictable only during periods when the correlation between market returns and variance innovations is highly negative. These results confirm the hypothesis that the VRP is a better measure of the monthly market risk premium when variance risk explains a higher proportion of market risk.

Table 9
Alternative specifications.
This table repeats the analyses of Table 2 for several alternative specifications. In Panel A, weighted least squares (WLS) is used in the first step regressions of the traditional approach. Alternatively, a shorter estimation period of five or seven years is used for the first step regressions. Then, the first seven years of data are dropped from the analysis. In Panel B, the results of the contemporaneous beta approach are provided after dropping the first seven years.

		VRP_N	$\text{VRP}_{\overline{P}}$	VRP_{PE}
Panel A: The traditional approach				
WLS, ten-year rolling	OOS-R^2	0.016	0.005	0.050
1993–2016	Wald	0.319	0.026	1.806
	p-value	(0.572)	(0.872)	(0.179)
Five-year rolling	OOS-R^2	−0.046	−0.039	0.015
1993–2016	Wald	0.805	0.664	0.001
	p-value	(0.370)	(0.415)	(0.978)
Seven-year rolling	OOS-R^2	−0.030	−0.029	0.018
1993–2016	Wald	0.053	0.144	0.563
	p-value	(0.818)	(0.704)	(0.453)
Expanding window	OOS-R^2	0.015	0.005	0.057
1998–2016	Wald	0.112	0.016	1.488
	p-value	(0.738)	(0.899)	(0.223)
Panel B: The contemporaneous beta approach (no intercept)				
One-month	OOS-R^2	0.073	0.090	0.093
WLS beta	Wald	5.309	8.728	5.621
1998–2016	p-value	(0.021)	(0.003)	(0.018)
One-month	OOS-R^2	0.066	0.083	0.085
OLS beta	Wald	4.736	8.728	6.158
1998–2016	p-value	(0.030)	(0.003)	(0.013)

6.2. Alternative specifications for the traditional approach

In a recent study, Johnson (2017) examines the OOS performance of a number of common return predictors and shows that using WLS as the first-stage regression improves OOS predictability. I study whether using WLS regressions as in the traditional manner makes the forecast comparable to the performance of the contemporaneous beta approach. The first part of Panel A in Table 9 shows the result. Comparing the results to Table 2, WLS slightly improves OOS predictability. The R^2s sometimes increase as much as 0.6% but also decrease for other measures, for example, if the VRP is measured using only end-of-month observations. However, even in this case, the Wald statistic slightly improves. Overall, despite a slight increase in OOS predictability, they are nowhere comparable to the forecasts of the contemporaneous beta approach.

The traditional approach of return forecasting requires an estimate for the slope and intercept to form an OOS forecast. If the sampling period for the first step regression is too short, the estimate may contain too much estimation error, so that it does not provide a good forecast. Using excessive data for the first step regression is also problematic if the predictive relation rapidly changes over time. The traditional approach of return forecasting may fail in either of these circumstances.

[13] The in-sample regression for this measure is only marginally significant with t-statistics of 1.24–2.32. The interactive coefficients are all highly statistically significant with t-statistics above 3.0.

To deal with the possibility that the underperformance of the traditional approach is driven by not selecting the optimal first-stage estimation interval, I consider five-year/seven-year rolling windows for the first-stage regression. The next part of Panel A shows the results. The panel shows that using a shorter sample for the first step regression makes the out-of-sample predictive performance even worse.

For the same reason, I also drop the first eight years from the sample entirely and consider a subsample of 1998–2016. The last part of Panel A of Table 9 summarizes the performance for the sample of 1998–2016. Panel B summarizes the result of the contemporaneous beta approach. The contemporaneous beta approach strictly outperforms traditional return forecasting. The OOS-R^2s for the traditional approach slightly increase for the post-1998 period, but there is an equivalent increase in predictive performance using the contemporaneous beta approach. These are times when the betas are more negative and when the fluctuations in betas are better identified.

7. Conclusion

It is well known that the market reacts negatively to unexpected market variance shocks. This negative relation between market returns and variance innovations implies that the market is subject to variance risk. Using option prices and high-frequency intraday trading data, we can easily gauge the variance risk premium of the market relatively accurately. In particular, previous research finds that this risk premium is useful in predicting short-horizon market returns.

This paper studies how the market's exposure to variance risk is related to the time-varying predictability of market returns by the VRP. First, this article shows that the slope that determines the contemporaneous relationship between market and variance risk resembles the relationship between the risk premium of the market and market variance. As a result, when the beta of the contemporaneous regression of market returns on changes in its variance is used as the predictive slope for the VRP, one-month market returns can be predicted in a statistically and economically significant manner, even out of sample. Second, the predictive power strongly depends on the degree of the contemporaneous correlation between returns and variance innovations. When correlations are highly negative, predictions can be made more accurately. This result holds both in sample and out of sample. Since the predicted strength of the leverage effect can be estimated ex ante, we can anticipate this predictive power. The combination of the contemporaneous beta and the VRP outperforms the average returns consistently over time, regardless of the strength of the asymmetry in the market.

Although the VRP is constructed from option prices on the index as well as index returns, its ability to predict future returns is not necessarily restricted to the equity index. At least, theoretically, the VRP should predict all returns that highly correlate with changes in stock market variance. Furthermore, while not directly shown, this paper suggests that the predictive beta may be related to the exposure of the asset to market variance risk.

Acknowledgments

This paper is based on the first chapter of my doctoral dissertation at USC Marshall School of Business. I thank my advisor Christopher Jones for his continuous encouragement, guidance, and support. I am grateful to Bill Schwert (the editor), an anonymous referee, Wayne Ferson, Alladeen Hameed, Larry Harris, Kris Jacobs, Scott Joslin, Anh Le, Juhani Linnainmaa, Ralitsa Petkova, Johan Sulaeman, Selale Tuzel, Nancy Xu, and Fernando Zapatero for helpful comments. I also thank seminar participants at Case Western Reserve University, City University of Hong Kong, National University of Singapore, Penn State University, University of Hong Kong, University of Southern California, AFBC, EFA, MFA, and Young Scholars Finance Consortium at Texas A&M. I appreciate support by NUS start-up grant (R-315-000-120-133). All errors are my own.

References

Andersen, T.G., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. J. Finance 52 (3), 975–1005.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Am. Stat. Assoc. 96 (453), 42–55.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Rev. Financial Stud. 20, 651–707.
Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. J. Finance 61 (1), 259–299.
Bali, T.G., Zhou, H., 2016. Risk, uncertainty, and expected returns. J. Financial Quant. Anal. 51 (3), 707–735.
Bandi, F., Reno, R., 2016. Price and volatility co-jumps. J. Financial Econ. 119 (1), 107–146.
Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. J. Economet. 183 (2), 181–192.
Bekaert, G., Hoerova, M., LoDuca, M., 2013. Risk, uncertainty and monetary policy. J. Monet. Econ. 60 (7), 771–788.
Black, F., 1976. Studies of stock price volatility changes. In: Proceedings of the 1976 Meetings of the American Statistical Association, pp. 177–181.
Bollerslev, T., Gibson, M., Zhou, H., 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. J. Economet. 160 (1), 235–245.
Bollerslev, T., Litvinova, J., Tauchen, G., 2006. Leverage and volatility feedback effects in high-frequency data. J. Financial Economet. 4 (3), 353–384.
Bollerslev, T., Marrone, J., Xu, L., Zhou, H., 2014. Stock return predictability and variance risk premia: statistical inference and international evidence. J. Financial Quant. Anal. 49, 633–661.
Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financial Stud. 22 (11), 4463–4492.
Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. J. Financial Econ. 118 (1), 113–134.
Busch, T., Christensen, B.J., Nielsen, M.O., 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. J. Economet. 160 (1), 48–57.
Campbell, J.Y., 1987. Stock returns and the term structure. J. Financial Econ. 18 (2), 373–399.
Campbell, J.Y., Shiller, R.J., 1988. Stock prices, earnings, and expected dividends. J. Finance 43 (3), 661–676.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financial Stud. 21 (4), 1509–1531.
Carr, P., Wu, L., 2009. Variance risk premiums. Rev. Financial Stud. 22 (3), 1311–1341.
Carr, P., Wu, L., 2016. Analyzing volatility risk and risk premium in option contracts: a new theory. J. Financial Econ. 120 (1), 1–20.

Chang, B.Y., Christoffersen, P., Jacobs, K., 2013. Market skewness risk and the cross section of stock returns. J. Financial Econ. 107 (1), 46–68.

Chen, H., Joslin, S., Ni, S.X., 2018. Demand for crash insurance, intermediary constraints, and risk premia in financial markets. Rev. Financial Stud. in press.

Chen, L., 2009. On the reversal of return and dividend growth predictability: A tale of two periods. J. Financial Econ. 92 (1), 128–151.

Chen, X., Ghysels, E., 2011. News–good or bad–and its impact on volatility predictions over multiple horizons. Rev. Financial Stud. 24 (1), 46–81.

Choi, H., Mueller, P., Vedolin, A., 2017. Bond variance risk premiums. Rev. Finance 21 (3), 987–1022.

Christie, A.A., 1982. The stochastic behavior of common stock variances. J. Financial Econ. 10 (4), 407–432.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financial Economet. 7 (2), 174–196.

Cujean, J., Hasler, M., 2017. Why does return predictability concentrate in bad times? J. Finance 72 (6), 2717–2758.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. J. Financial Econ. 106 (1), 157–181.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263.

Drechsler, I., Yaron, A., 2011. What's vol got to do with it. Rev. Financial Stud. 24 (1), 1–45.

Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. J. Financial Econ. 22 (1), 3–25.

Fama, E.F., French, K.R., 1988. Permanent and temporary components of stock prices. J. Polit. Econ. 96 (2), 246–273.

Ferson, W.E., Sarkissian, S., Simin, T.T., 2003. Spurious regressions in financial economics? J. Finance 58 (4), 1393–1413.

Feunou, B., Jahan-Parvar, M.R., Okou, C., 2017. Downside variance risk premium. J. Financial Economet, 16 (3), 341–383.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. J. Financial Econ. 19 (1), 3–29.

Garcia, D., 2013. Sentiment during recessions. J. Finance 68 (3), 1267–1300.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.

Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. J. Finance 48 (2), 663–679.

Gonzalez-Urteaga, A., Rubio, G., 2016. The cross-sectional variation of volatility risk premia. J. Financial Econ. 119 (2), 353–370.

Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financial Stud. 21 (4), 1455–1508.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econ. Stat. 24 (2), 127–161.

Henkel, S.J., Martin, J.S., Nardari, F., 2011. Time-varying short-horizon predictability. J. Financial Econ. 99 (3), 560–580.

Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Rev. Financial Stud. 5 (3), 357–386.

Johannes, M., Korteweg, A., Polson, N., 2014. Sequential learning, predictability, and optimal portfolio returns. J. Finance 69 (2), 611–644.

Johnson, T. L., 2017. A Fresh Look at Return Predictability Using a More Efficient Estimator. Unpublished working paper. University of Texas at Austin.

Jones, C.S., Tuzel, S., 2013. New orders and asset prices. Rev. Financial Stud. 26 (1), 115–157.

Keim, D.B., Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. J. Financial Econ. 17 (2), 357–390.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financial Stud. 27, 2841–2871.

Lettau, M., Ludvigsen, S.C., 2001. Consumption, aggregate wealth, and expected stock returns. J. Finance 56 (3), 815–849.

Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Rev. Financial Stud. 21 (4), 1607–1652.

Londono, J. M., 2014. The Variance Risk Premium Around the World. Unpublished working paper Federal Reserve Board of Governors.

Londono, J.M., Zhou, H., 2017. Variance risk premiums and the forward premium puzzle. J. Financial Econ. 124 (2), 414–440.

Lustig, H., Roussanov, N., Verdelhan, A., 2014. Countercyclical currency risk premia. J. Financial Econ. 111 (3), 527–553.

Martin, I., Wagner, C., 2016. What is the expected return on the market? Q. J. Econ. 132 (1), 367–433.

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41 (5), 867–887.

Pastor, L., Stambaugh, R., 2009. Predictive systems: Living with imperfect predictors. J. Finance 64 (4), 1583–1628.

Pesaran, M.H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. J. Finance 50 (4), 1201–1228.

Pindyck, R.S., 1984. Risk, inflation, and the stock market. Am. Econ. Rev. 74 (3), 335–351.

Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. J. Financial Econ. 22 (1), 27–59.

Rapach, D.E., Ringgenberg, M.C., Zhou, G., 2016. Short interest and aggregate stock returns. J. Financial Econ. 121 (1), 46–65.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financial Stud. 23 (2), 821–862.

Stambaugh, R.F., 1999. Predictive regressions. J. Financial Econ. 54 (3), 375–421.

Todorov, V., 2010. Variance risk-premium dynamics: the role of jumps. Rev. Financial Stud. 23 (1), 345–383.

Wang, H., Zhou, H., Zhou, Y., 2013. Credit default swap spreads and variance risk premia. J. Bank. Finance 37 (10), 3733–3746.

Zhang, L., Mykland, P., Ait-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. 100 (472), 1394–1411.