The Information Content of the Implied Volatility Term Structure on Future Returns

ABSTRACT

We adopt the Heston (1993) stochastic volatility (SV) model framework to examine the theoretical relation between the term structure of implied volatility and the expected excess returns of underlying assets. Three alternative approaches are adopted for our compilation of the variables representing the information on the squared VIX level and term structure in support of our empirical investigation of the information content of the level and term structure variables on future excess returns in the S&P 500 index. Our empirical results provide support for the important role of the term structure in the determination of future excess returns, with such predictive power being discernible for various horizons. Overall, the information content of the term structure variable is found to be significant, and indeed, a strong complement to that of the level variable. In particular, due to the mean-reversion behavior of volatility, the information in the term structure of implied volatility is found to be very effective in the prediction of shorter-term excess returns.

Keywords: VIX term structure, Predictability, S&P 500 index returns.

JEL Classification: G13, G14
1. INTRODUCTION

According to the methodology proposed by the CBOE for the implementation of the model-free implied volatility formula,¹ the computation of the implied volatility index (VIX) is based upon consideration of all of the available market prices of the S&P 500 index options. Such an approach facilitates the approximation of the expected aggregate volatility of the S&P 500 index during the subsequent 30 calendar-day period,² and indeed, this method has been used not only as a measure of sentiment but also as an instrument for timing the market, particularly in the aftermath of the subprime mortgage crisis.

Following Whaley (2000), who proposed the use of the VIX as an effective fear indicator, Giot (2005) identified a strongly negative correlation between contemporaneous changes and future market index returns, along with a positive correlation between such future returns and current levels of the implied volatility indices. Both Guo and Whitelaw (2006) and Banerjee, Doran, and Peterson (2007) reported similar findings; for example, Banerjee et al. (2007) derived the theoretical relation between the level and innovation of the VIX and future returns by adopting Heston’s (1993) stochastic volatility (SV) model. Their results provided empirical

¹ The CBOE’s revision of the methodology of the VIX formula was based largely on the results of Carr and Madan (1998) and Demeterfi et al. (1999). As regards theoretical fundamentals, Britten-Jones and Neuberger (2000) derived model-free implied volatility under the diffusion assumption, with Jiang and Tian (2005) subsequently extending the model to the jump diffusion assumption.

² When it was originally introduced in 1993, the VIX was compiled from the implied volatility of eight S&P 100 index options, comprising near at-the-money and nearby and second nearby calls and puts; however, ever since 2003, the VIX has been calculated from the prices of S&P 500 index options using a model-free formula with almost all of the available contracts, that is, with a wide range of strike prices.
support for the predictive ability of the VIX-related variables on future portfolio returns.

Because the current version of the VIX is compiled for each maturity period with the incorporation of almost all of the available contracts, it should be more informative than the old version when used to investigate its predictive ability on future equity returns; however, the VIX is dependent on maturity periods, with the 30-day version being the most frequently used. This dependence on maturity periods gives rise to the interesting and important question of whether the VIX term structure contains any useful information for potential use in the forecasting of returns.³

Furthermore, because it is well known that volatility has some special stylized facts, such as clustering and mean-reversion, the relative positions of the VIX levels for different maturity periods may imply the expectations of market participants on market volatility and, thus, on changes in the S&P 500 index due to the mean-variance relation in the conventional theory of the capital asset pricing model. Therefore, this study contributes to the extant related literature by comprehensively investigating whether the VIX term structure contains any useful information on future returns in the S&P 500 index.

Based on Heston’s (1993) SV model framework, we refine the theoretical work of Banerjee et al. (2007) by deriving a theoretical model that reveals a positive relation between expected excess returns in the S&P 500 index and both the squared VIX levels and the

³ See, for example, Egloff, Leippold, and Wu (2010), Bakshi, Panayotov, and Skoulakis (2011), Duan and Yeh (2011), Luo and Zhang (2012b), and Feunou, Fontaine, Taamouti, and Tedongap (2014).
difference between forward and current squared VIX levels. Because the forward squared VIX level can be computed from two squared VIX values with different horizons, it can be regarded as a proxy for the VIX term structure. Hence, this model provides theoretical fundamentals for the potential of the VIX term structure with regard to the prediction of excess returns in the S&P 500 index.

We propose three alternative empirical methods for compiling the variables representing the information in both the squared VIX level and its term structure to investigate whether the information content of the term structure contributes to the forecasting of future excess returns in the S&P 500 index. First, adhering closely to the theoretical model, we use the 30-day squared current VIX level and the difference between the 30-day forward squared VIX and the 30-day squared current VIX level as the respective level and term structure variables. Second, we run a principal component analysis (PCA) of the squared VIX levels under various time horizons to generate the first and second components for our investigation; the first and second components represent the level and slope of the VIX term structure, respectively. Third, we employ the two-factor SV model proposed by Egloff et al. (2010) to transform the maturity-dependent squared VIX values into maturity-independent instantaneous variance and stochastic central tendency. We then use the maturity-independent instantaneous variance along with the difference between the maturity-independent instantaneous variance and the stochastic central tendency as the proxies for the squared VIX level and term structure, respectively.
Our findings reveal that the information content of the term structure plays an important role in the prediction of excess returns regardless of which approach is used for the compilation of the variables; furthermore, this predictive power is also discernible for the excess returns under various horizons. When comparing the incremental contribution provided by the level and term structure variables with regard to the effective prediction of returns, we find that the information content of the squared VIX term structure is significant and, indeed, a strong complement to that of the squared VIX level. In particular, as a result of the mean-reversion behavior of volatility, the term structure variable is more informative over shorter-term prediction horizons. Our empirical results are also insensitive to the use of either overlapping or non-overlapping data.

Attempts have been made in recent studies to estimate specific price dynamics by incorporating the VIX term structure under a two-factor volatility framework;\(^4\) however, the information content of the VIX term structure on future excess returns in the S&P 500 index has yet to attract any significant attention. Cochrane and Piazzesi (2005) provided an effective approach for extracting the information content of the interest rate term structure for the prediction of excess bond returns. Bakshi, Panayotov, and Skoulakis (2011) and Luo and Zhang (2012b) subsequently applied the approach to their empirical investigations of the relation between forward variances and future excess returns. Johnson (2016) investigates the

---

\(^4\) Examples include Egloff et al. (2010), Duan and Yeh (2011), and Luo and Zhang (2012a).
predictability of the VIX term structure with PCA for the variance-related asset returns. However, neither study offered any theoretical fundamentals for the prediction of these relations.

Furthermore, although Feunou et al. (2014) derived a state-dependent relation between S&P 500 excess returns and the VIX for different maturity periods, the direction of the impact remained somewhat ambiguous. Our study therefore contributes to the literature by not only proposing a theoretical model but also providing empirical evidence in support of the important role of the information in the VIX term structure on the prediction of future excess returns in the S&P 500 index.

The remainder of this paper is organized as follows. Section 2 provides the theoretical fundamentals adopted for our study. Section 3 describes the data and empirical methodologies used for our empirical analyses. Section 4 presents the empirical results and robustness tests. Finally, Section 5 discusses the conclusions drawn from this study.

2. THEORETICAL FUNDAMENTALS

This study proposes a model to examine the relation between future excess returns and the term structure of implied volatility by referring to Banerjee et al.’s (2007) model. Following the stochastic volatility (SV) setting outlined in Heston (1993), we specify the dynamics of the asset prices, $S_t$, and the variance in the asset prices, $V_t$, under the real-world $P$ measure as
\[ dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^p, \quad (1) \]
\[ dV_t = \kappa (\theta - V_t) dt + \lambda \sqrt{V_t} dW_t^p + \sigma \sqrt{V_t} dZ_t^p, \quad (2) \]

where \( \text{cov}(dW_t^p, dZ_t^p) = \rho dt \); \( \mu = r_f + \lambda V_t \) is the expected return; \( r_f \) is the risk-free rate; \( \kappa \) is the volatility mean-reversion speed; \( \theta \) is the long-run variance level; \( \sigma \) is the volatility of the volatility; \( \rho \) is the correlation between the price and variance innovations (i.e., \( W_t^p \) and \( Z_t^p \)); \( \lambda \) (\( \lambda_v \)) is the market price of the price (variance) risk. Using the Girsanov theorem to transform the real-world processes into their risk-neutral \( Q \)-measure equivalents, we obtain the following \( Q \)-measure dynamics:

\[ dS_t = r_f S_t dt + \sqrt{V_t} S_t dW_t^Q, \quad (3) \]
\[ dV_t = \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dZ_t^Q, \quad (4) \]

where \( \text{cov}(dW_t^Q, dZ_t^Q) = \rho dt \), where \( W_t^Q \) and \( Z_t^Q \) are the \( Q \)-measure price and variance innovations, respectively.

By discretizing the \( P \)-measure dynamics in Equations (1) and (2) and then replacing \( V_t \) with the realized variance, \( \sigma_{RV,t}^2 \), due to the \( P \)-measure, we obtain

\[ S_{t+1} - S_t = (r_f + \lambda \sigma_{RV,t}^2) S_t + \sigma_{RV,t} S_t \epsilon_t \quad (5) \]
\[ \sigma_{RV,t+1}^2 - \sigma_{RV,t}^2 = \kappa (\theta - \sigma_{RV,t}^2) + \lambda \sigma_{RV,t}^2 \sigma_{RV,t} + \sigma \sigma_{RV,t} (\rho \epsilon_t + \sqrt{1 - \rho^2} \epsilon_t^\nu), \quad (6) \]

where \( \epsilon_t \) and \( \epsilon_t^\nu \) are the discrete-version price variance innovations under the \( P \)-measure and \( \epsilon_t \) and \( \epsilon_t^\nu \) are uncorrelated (i.e., \( \text{Cov}(\epsilon_t, \epsilon_t^\nu) = 0 \)).

Similarly, by discretizing the \( Q \)-measure volatility dynamic in Equation (4) or applying...
the Girsanov theorem to Equation (6) and then replacing $V_t$ with the implied variance, $\sigma^2_{IV,t}$, due to the $Q$-measure, we obtain

$$\sigma^2_{IV,t+1} - \sigma^2_{IV,t} = \kappa (\theta - \sigma^2_{IV,t}) + \sigma_V \sigma_{IV,t} (\rho \varepsilon_t^* + \sqrt{1 - \rho^2} \varepsilon_t^{V*}),$$

where $\varepsilon_t^{V*}$ and $\varepsilon_t^{V}$ are the discrete-version price variance innovations under the $Q$-measure and they are uncorrelated (i.e., $\text{Cov}(\varepsilon_t^{V*}, \varepsilon_t^{V}) = 0$).

To connect the price risk with the variance risk, we assume that

$$\lambda = -\delta \lambda_V,$$

where $\delta$ is restricted to be positive.

While the price risk premiums, $\lambda$, are widely known to be positive, in some related studies, such as Carr and Wu (2009), the variance risk premiums, $\lambda_V$, are strongly negative; our finding of a proportional relation between the two risk premiums is therefore consistent with the findings in the extant literature.\(^5\)

By applying Equation (8) to Equation (6) and then taking the difference with Equation (7), we obtain

$$\lambda \sigma^2_{RV,t} = \delta (\Delta \sigma^2_{IV,t} - \Delta \sigma^2_{RV,t}) + \kappa \delta (\sigma^2_{IV,t} - \sigma^2_{RV,t}) + \delta g_1 (\varepsilon_t, \varepsilon_t^{V}, \varepsilon_t^{V*}, \varepsilon_t^{V*}),$$

where

\(^5\) Our assumption on the relation between the price and volatility risks differs from that of Banerjee et al. (2007), in which $\lambda = \lambda_v + \delta$, where $\delta$ is restricted to be greater than the absolute value of the variance risk premium. Our assumption is more general, and because it imposes less restriction on the $\delta$ parameter, it always results a positive price and negative variance risk premiums.
\[ g_1(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t) = \sigma \left( \sigma_{RV,t} \left( \rho \epsilon_t + \sqrt{1 - \rho^2} \epsilon^*_t \right) - \sigma_{IV,t} \left( \rho \epsilon_t + \sqrt{1 - \rho^2} \epsilon^{V*}_t \right) \right) \]

and \( \Delta \sigma^2_t = \sigma^2_{t+1} - \sigma^2_t \).

Equations (5) and (9) then jointly result in the following relation between future excess returns and realized and implied volatility levels:

\[ r_{t,t+1} - r_f = \delta (\Delta \sigma^2_{IV,t} - \Delta \sigma^2_{RV,t}) + \kappa \delta (\sigma^2_{IV,t} - \sigma^2_{RV,t}) + g_2(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t), \tag{10} \]

where

\[ r_{t,t+1} = \frac{s_{t+1} - s_t}{s_t} \]

and \( g_2(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t) = \sigma_{RV,t} s_t \epsilon_t - g_1(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t) \).

Because the literature already well documents that the implied volatility computed from option prices is an efficient, albeit upwardly biased, forecast of realized volatility\(^6\) and highly self-correlated, we assume that

\[ \sigma^2_{RV,t} = \alpha + \Psi \sigma^2_{IV,t}, \tag{11} \]

where \( \Psi \) should be less than 1, due to the upwardly biased prediction.

By applying Equation (11) to Equation (10), we obtain the relation between future excess returns and implied volatility levels across time points as

\[ r_{t,t+1} - r_f = \delta (1 - \Psi) \Delta \sigma^2_{IV,t} + \kappa \delta (1 - \Psi) \sigma^2_{IV,t} + g_3(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t), \tag{12} \]

where \( g_3(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t) = g_2(\epsilon_t, \epsilon^*_t, \epsilon^*_t, \epsilon^{V*}_t) - \alpha \kappa \delta. \)

Taking the conditional expectation for Equation (12), we derive the relation between the

---

\(^6\) See, for example, Christensen and Prabhala (1998), Fleming (1998), Blair, Poon, and Taylor (2001), and Jiang and Tian (2005).
expected future excess returns and the expected future and current implied volatility levels as

\[ E_t(r_{t,t+1} - r_f) = A_1\left(E_t(\sigma^2_{IV,t+1}) - \sigma^2_{IV,t}\right) + A_2\sigma^2_{IV,t} + K, \]  \hspace{1cm} (13)

where \( A_1 = \delta(1 - \Psi), \) \( A_2 = \kappa \delta(1 - \Psi), \) and \( K = E_t\left(g_3(\varepsilon_t, \varepsilon^V_t, \varepsilon^*_t, \varepsilon^{V*}_t)\right). \) Because both \( \delta \) and \( \kappa \) are positive, and \( \Psi \) is less than 1, both \( A_1 \) and \( A_2 \) are also positive.

Equation (13) provides a theoretical fundamental for the potential predictive ability on excess returns arising not only from the current implied volatility level but also from its term structure, which contains the expectation on future implied volatility. The positive sign on \( A_2 \) indicates a positive association between expected excess returns and risk-neutral expected variance, which is consistent with the positive return–risk relation stated in the conventional asset pricing models. The positive sign of \( A_1 \) indicates a positive linkage between expected excess returns and the expectation on changes in expected variance, which implies that the expectation of an upward change in implied variance may drive future excess returns in the same direction.

3. DATA AND METHOD

3.1. Data

The daily S&P 500 index levels and the prices of the options written on the index are obtained from OptionMetrics. Our sample period (2 January 1998 to 31 August 2012) provides a rather rich data period because it includes both bull and bear regimes. The VIX levels for all of the
available time horizons are calculated based on the prices of S&P 500 index options and their corresponding time to maturity. The bond data for analysis in this study are obtained from the Federal Reserve Bank of St. Louis,\textsuperscript{7} and the high-frequency index levels are obtained from OlsenData.

We construct the squared VIX for several standardized horizons, comprised of 30, 60, 90, 180, 270, and 360 calendar days, following the method and the interpolation process adopted by the CBOE.\textsuperscript{8} The squared VIX value serves as the measure for the risk-neutral expected variance. Table 1 shows that the excess returns of the S&P 500 index are more correlated with $DVIX_{7,30}^2$ than $VIX_{7,30}^2$, which may be a signal of the potential predictive ability of the VIX term structure with regard to returns. The highly negative correlation (-0.62) between $DVIX_{7,30}^2$ and $VIX_{7,30}^2$ arises as a result of the mean reversion property of volatility.

\textless TABLE 1 ABOUT HERE\textgreater

Table 2 presents the correlation matrix of the squared VIX values across various maturity periods. Because all values are clearly highly correlated, it is inappropriate to include them simultaneously within a regression model, which lends additional support to the

\textsuperscript{7} See http://research.stlouisfed.org/ for more details.
\textsuperscript{8} See the VIX white paper “The CBOE Volatility Index—VIX” (http://www.cboe.com/micro/vix/vixwhite.pdf) for more details.
appropriateness of our approach of extracting the term structure information from the slope or through PCA.\footnote{Feunou et al. (2014) and Johnson (2016) also used PCA to extract the information content of the VIX term structure.}

As suggested in many prior studies, several option-implied variables (e.g., variance risk premium and model-free skewness and kurtosis) and macroeconomic variables (e.g., interest rate term spread and default spread) are also informative with regard to future stock returns.\footnote{Examples include Ang and Bekaert (2007), Bollerslev, Tauchen, and Zhou (2009), Vilkov and Xiao (2013), Conrad, Dittmar, and Ghysel (2013), Chang, Christoffersen, and Jacobs (2013) and Kozhan, Neuberger, and Schneider (2013).} Thus, we follow Bollerslev et al. (2009) to select three factors as the control variables in our regression analysis: (i) variance risk premium ($VRP$), which is defined as the difference between the de-annualized squared VIX level and the realized variance compiled from the high-frequency five-minute transaction prices;\footnote{We follow Bollerslev et al. (2009) to use intra-day data to construct the realized monthly variance. As discussed in the prior studies, including Andersen, Bollerslev, Diebold, and Labys (2001) and Hansen and Lunde (2006), the selection of the sampling frequency is the trade-off between data continuity and market microstructure noises. Five minutes is the most frequently adopted frequency for the calculation of stock realized volatility.} (ii) default spread ($DFS$), which is defined as the difference between Moody’s BAA and AAA bond yield indices; and (iii) term spread ($TMS$), which is defined as the difference between the ten-year and three-year Treasury yields.

\begin{table}[h]
\centering
\caption{Table 2: Control Variables Used in the Regression Analysis}
\begin{tabular}{|c|c|}
\hline
Variable & Definition \\
\hline
$VRP$ & Variance Risk Premium, defined as the difference between the de-annualized squared VIX level and the realized variance compiled from the high-frequency five-minute transaction prices. \\
$DFS$ & Default Spread, defined as the difference between Moody’s BAA and AAA bond yield indices. \\
$TMS$ & Term Spread, defined as the difference between the ten-year and three-year Treasury yields. \\
\hline
\end{tabular}
\end{table}
We use additional controls for the influence of model-free skewness \( (SKEW) \) and kurtosis \( (KURT) \), which are calculated based on the method proposed by Bakshi, Kapadia, and Madan (2003). Because our empirical analysis focuses on the S&P 500 index, we construct the VIX using CBOE’s methodology as the implied volatility index. The appendix provides details on the computation of the VIX and the construction of the VIX term structure.

3.2. Method

We propose three alternative approaches to the incorporation of the information implied in the VIX term structure for the prediction of future excess returns.

3.2.1. Forward implied variance

Our first approach adheres closely to our theoretical model. \( E_t(\sigma_{IV,t+1}^2) \), by definition, refers to the expectation at time \( t \) of the implied variance at time \( t+1 \), which gauges the level of volatility during the period from time \( t+1 \) to \( t+2 \). Conceptually, this provides the forward variance for the period from time \( t+1 \) to \( t+2 \); therefore, we replace \( E_t(\sigma_{IV,t+1}^2) \) with the forward implied variance, \( 2VIX_{t,2}^2 - VIX_{t,1}^2 \), which is computed from the squared VIX term structure.

Given that one month (30 calendar days) is commonly used as the time horizon in the VIX index, we also set the time unit as one month (30 days) and then rewrite Equation (13) as
\[ E_t(r_{t,t+30} - r_f) = A_1' * DVIX^2_{t,30} + A_2 * VIX^2_{t,30} + K, \]  

(14)

where \( DVIX^2_{t,30} = FVIX^2_{t,(30,60)} - VIX^2_{t,30} = 2VIX^2_{t,60} - VIX^2_{t,30} - VIX^2_{t,60} = 2(VIX^2_{t,60} - VIX^2_{t,30}) \).

We run the following regression to examine whether the information implied in the VIX term structure can predict excess returns:

\[ ER_{t,t+h} = \alpha + \beta_1 VIX^2_{t,30} + \beta_2 DVIX^2_{t,30} + \epsilon_t, \]  

(15)

where \( ER_{t,t+h} \) refers to the excess return for the period from time \( t \) to \( t+h \), which is defined as the return on the S&P 500 index minus the three-month T-bill rate. According to our theoretical model, \( \beta_1 \) and \( \beta_2 \), which, respectively, represent the predictive contribution of the implied variance level and the term structure with regard to excess returns, are both expected to be positive.

3.2.2. The first and second principal components

Cochrane and Piazzesi (2005) indicate that in the bond markets the first three components, generated from PCA (i.e., level, slope, and curvature of the interest rate term structure) provide abundant information on excess bond returns. Furthermore, Feunou et al. (2014) apply PCA to extract the systematic factors across the term structure of option-implied variance and empirically and show that the first two components are sufficient to explain the changes in future S&P 500 index returns.

Because \( DVIX^2_{t,30} = 2(VIX^2_{t,60} - VIX^2_{t,30}) \) is a conceptual measure of the slope of the implied
variance term structure for the period from 30 to 60 days, it would seem natural to question whether the slope of the implied variance term structure also contains useful information on the future excess returns of the underlying asset.

Following the approaches adopted in several studies within the extant literature, we apply PCA to the squared VIX across all of the available maturity periods and take the first and second principal components, $PC_1$ and $PC_2$, as the information proxy of the implied variance term structure. These two components essentially represent the level and slope of the implied variance term structure, which are the conceptual equivalents of $VIX_{t, 30}^2$ and $DVIX_{t, 30}^2$ respectively. Therefore we amend the regression model to

$$ER_{t,t+h} = \alpha + \beta_1 PC_1 + \beta_2 PC_2 + \epsilon_t. \tag{16}$$

### 3.2.3. Two-factor stochastic volatility framework proxies

Egloff et al. (2010) extended Heston’s (1993) SV model to allow the central tendency of the variance to be another stochastic process; thus, they revised the decomposition of the squared VIX to

$$VIX_{t, \tau}^2 = \omega_1 V_t + \omega_2 m_t + (1 - \omega_1 - \omega_2) \theta_{m_t} \tag{17}$$

where

$$\omega_1 = \frac{1 - e^{-\kappa_{\tau}}}{\kappa_{\tau}}, \quad \omega_2 = \frac{1 + \frac{k_m}{\kappa_{m}} e^{-\kappa_{\tau}} - \frac{\kappa_{\nu}}{\kappa_{\nu} - \kappa_{m}} e^{-\kappa_{m} \tau}}{\kappa_{m} \tau}.$$

12 From our analysis, $PC_1$ and $PC_2$ are capable of explaining almost 95% of the variation in the VIX term structure.
Consequently, the squared VIX indices for various maturity periods can be transformed to the maturity-independent instantaneous variance, $V_t$, the stochastic central tendency, $m_t$, and the long-run mean, $\theta_m$.\(^{13}\)

Researchers can quite straightforwardly adopt instantaneous variance as the proxy for the level of the VIX term structure. Given the mean-reversion property of volatility, the relative relation between the instantaneous variance and the stochastic central tendency, $m_t - V_t$, may reveal the direction in which the VIX level is likely to move, which is the analogue of the slope of the VIX term structure.

We use the efficient iterative two-step procedure suggested by Christoffersen, Heston, and Jacob (2009) to estimate the parameters and to generate the time series of $V_t$ and $m_t$. The two-stage procedure is implemented as follows.

**Step 1**: We solve the following optimization in order to estimate the time series of $(\hat{V}_t, \hat{m}_t)$, $t = 1,2,\ldots, T$.

$$\left(\hat{V}_t, \hat{m}_t\right) = \text{argmin} \sum_{j=1}^{n_t} \left(VIX^{Mkt}_{t,\tau_j} - VIX^2_{t,\tau_j}\right)^2, t = 1,2,\ldots, T, \quad (18)$$

where $VIX^{Mkt}_{t,\tau_j}$ and $VIX_{t,\tau_j}$, respectively, denote the market and theoretical VIX levels for the maturity $\tau_j$ at time $t$.

**Step 2**: We collect the $(V_t, m_t)$ time series to estimate the $(\hat{\theta}_m, \hat{\kappa}_t, \hat{\kappa}_m)$ parameter by implementing the following minimization:

\(^{13}\) Luo and Zhang (2012a) also proposed a similar decomposition of the squared VIX under a framework with an independent stochastic long-run mean variance.
\[
(\hat{\theta}_m, \hat{k}_r, \hat{\kappa}_m) = \arg \min \sum_{t=1}^{T} \sum_{j=1}^{n_t} \left( VIX_{t,j}^{Mkt^2} - VIX_{t,j}^{2} \right)^2. \quad (19)
\]

The iteration procedure between Steps 1 and 2 is then carried out until the convergence criterion in the objective function of Step 2 is reached. The prediction regression is therefore revised to

\[
ER_{t,t+h} = \alpha + \beta_1 V_t + \beta_2 (m_t - V_t) + \epsilon_t. \quad (20)
\]

4. EMPIRICAL RESULTS

Our empirical analysis begins with an investigation into the information content of the VIX term structure for the subsequent-period excess returns with the variables compiled from the three different approaches described in Section 3. If predictive power is discernible, we then explore how long such power may persist. We subsequently determine whether any profitable trading strategy can be formed based on the predictive power of the VIX term structure. Finally, we carry out tests to verify the robustness of the results.

4.1. Forward Implied Variance Predictions

The \(DVIX_{t,30}^2\) variable is compiled from the forward variance implied in the VIX term structure. Given that this variable is perfectly in line with our theoretical model, we first run the regression model specified in Equation (15). To avoid the collinearity between \(DVIX_{t,30}^2\) and \(VIX_{t,30}^2\), we use the residual of \(DVIX_{t,30}^2\) regressed on \(VIX_{t,30}^2\) instead of \(DVIX_{t,30}^2\) in the following
model. Based on the control variables (see Section 3.2), the regression model is respecified as

\[
ER_{t,t+30} = \alpha + \beta_1 VIX_{t,30}^2 + \beta_2 DVIX_{t,30}^2 + \beta_3 VRP_t + \beta_4 SKEW_t + \\
+ \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon_t.
\]  

(21)

Table 3 reports the results on the full model and the various restricted models. Models 1 and 2 show that the coefficient of \(DVIX_{t,30}^2\), \(\beta_1\), is significantly positive whereas the coefficient of \(VIX_{t,30}^2\), \(\beta_1\), is positive but insignificant. When \(VIX_{t,30}^2\) and \(DVIX_{t,30}^2\) are run with the control variables in Models 4 and 5, respectively, \(DVIX_{t,30}^2\) is more informative than \(VIX_{t,30}^2\) on future excess returns. Finally, Model 6 shows that both variables are significantly positive for the full model.

Overall, the results show that the information implied in the squared VIX term structure plays an important role in the prediction of excess returns in the S&P 500 index, although the squared VIX level is also informative to some extent. The important role of \(DVIX_{t,30}^2\) also gains support from the higher incremental \(R^2\) values in those models in which \(DVIX_{t,30}^2\) is included.

In sum, the positive signs of \(\beta_1\) and \(\beta_2\) are consistent with our theoretical predictions, although they are not always significant. A comparison of the relative contributions of \(VIX_{t,30}^2\) and \(DVIX_{t,30}^2\) to the prediction of future excess returns shows that the latter is more informative than the former and that neither can completely replace the other. Therefore, in addition to the
squared VIX level, the squared VIX term structure also contains significant information of relevance to the prediction of excess returns in the S&P 500 index.

4.2. First and Second Principal Component Predictions

Given that the first and second principal components of the squared VIX term structure, $PC_1$ and $PC_2$, represent its level and slope, we use the PCA approach for the squared VIX values, under various time horizons, as the means of generating $PC_1$ and $PC_2$, and then run the model specified in Equation (16). The regression model with control variables included is respecified as

$$ER_{t,t+30} = \alpha + \beta_1 PC_{1t} + \beta_2 PC_{2t} + \beta_3 VR_{t} + \beta_4 SKW_{t} + \beta_5 KURT_{t} + \beta_6 TMS_{t} + \beta_7 DFS_{t} + \epsilon_t.$$  \hspace{1cm} (22)

Table 4 provides the results on the full model and the various restricted models. The first and second principal component results are generally in line with those obtained from forward implied variance, with both $PC_1$ and $PC_2$ positively related to future excess returns, which is consistent with our predictions. When examining those models without the control variables (Models 1–3), $PC_2$ is more informative than $PC_1$, because $\beta_2$ has greater significance than $\beta_1$. However, when the control variables are included, $\beta_1$ (Model 4) is significantly positive at the 5% level, and $\beta_2$ (Model 5) is significantly positive at the 10% level. Both $\beta_1$ and $\beta_2$ remain significant at level 5% in for the full model (Model 6).
Overall, these results show that, in addition to the squared VIX level, the slope of the squared VIX term structure contributes to the prediction of excess returns in the S&P 500 index although PC2 can explain only 5.5% of the variation in the squared VIX term structure whereas PC1 explains 93.11%. Our empirical findings provide strong support for the informativeness of the slope factor of the VIX term structure on future excess returns in the S&P 500 index.

4.3. Two-Factor Stochastic Volatility Framework Proxy Predictions

Because the two-factor stochastic volatility model provides an alternative instrument for the transformation of the maturity-dependent VIX values to maturity-independent instantaneous variance and its stochastic central tendency, we generate the time series of \((V_t, m_t)\) and then run the regression model specified in Equation (20). The regression model with the control variables included is respecified as

\[
ER_{t,t+30} = \alpha + \beta_1 V_t + \beta_2 (m_t - V_t) + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon_t.
\]

Table 5 provides the results on the full model and the various restricted models. The findings are generally consistent with those of the previous two alternative approaches. In
particular, both $\beta_1$ and $\beta_2$ are significantly positive at the 1% level in the full model. Therefore, in addition to the instantaneous variance level, its position relative to the stochastic central tendency plays a significant role in terms of predicting excess returns.

Regarding the incremental contribution to such predictions, $\beta_2$ is clearly more informative than $\beta_1$, because $\beta_2$ is significantly positive in most of the models. In sum, the findings from our third alternative approach confirm the importance of the information provided by the squared VIX term structure with regard to the prediction of excess returns in the S&P 500 index.

4.4. Predictions across Time Horizons

Thus far we focus on a one-month prediction horizon to match the maturity of the options used to compute the VIX index. Because our alternative approaches provide consistent findings on the important role of the squared VIX term structure in the prediction of excess returns, we investigate just how far forward the excess returns can be predicted. To investigate this intriguing question, we employ various horizons (2, 3, 6, 9, and 12 months) of the excess returns as the dependent variables in the regression models. Table 6 provides the results from the three alternative approaches.
Models 1 and 3 of Table 6 show that for all horizons both $\beta_1$ and $\beta_2$ are significantly positive at least at the 10% level. These findings indicate that the squared VIX level and the term structure are not only informative for shorter excess return horizons but also for longer horizons. Because this predictive power can be satisfactory for horizons of up to a full year, the relative contributions of the squared VIX level and the term structure to the prediction of excess returns across such horizons is clearly of interest.

To facilitate an exploration of the way in which the term structure factor incrementally contributes to such prediction, we compare the $R^2$ values for both the full model and the restricted model (where $\beta_2 = 0$). Table 6 provides the $R^2$ values of the restricted model for various prediction horizons and the percentage increase in $R^2$ as a result of the inclusion of the term structure variables. Regardless of the approach adopted for the compilation of the variables, the percentage contribution made by the term structure variable is most notable for the one-month prediction horizon, followed by a general decline over longer prediction horizons. For example, the value of $R^2$ for the full model (Panel C) increases by a factor of about two from 0.0448 to 0.0142 as compared to the model that includes $V_t$ and the control variables.

When using PCA as an alternative instrument for the compilation of the variables, the
contribution of the squared VIX term structure is not as impressive as that of the squared VIX level. For example, the $\beta_2$ coefficient is significantly positive at the 1% level only in the six-month case. Nevertheless, the sign remains positive across all prediction horizons, and the conclusions drawn from the $R^2$ values are consistent with those drawn from the two alternative approaches.

Overall, whereas both the VIX level and the VIX term structure are informative with regard to future excess returns in the S&P 500 index across all of the horizons investigated, the VIX term structure is particularly informative for shorter-horizon excess returns.

4.5. Trading Strategy Tests

According to the regression results previously discussed, information derived from the squared VIX term structure and the squared VIX level can be used to determine future excess returns in the S&P 500 index. However, our regression analysis is based on an in-sample framework. Therefore, as a guide for trading strategies, we take our analysis a step further to investigate the out-of-the-sample performance of the level and term structure based on predictions generated from regression models using historical data. Specifically, we run the models with the variables representing the squared VIX level and the squared VIX term structure with all control variables included to determine whether trading strategies using the information derived from the squared VIX term structure are more profitable than those using the squared VIX level only.
The trading strategies include (i) setting up a long position with a single unit asset when the forecast exceeds a critical value; (ii) setting up a short position with a single unit asset when the forecast is below the negative critical value; (iii) doing nothing when the forecast is within the positive and negative critical values; and (iv) the transaction cost is set to be 10 basis points for each trade. The critical values range from 0.01 to 0.07 (=7%), and the prediction horizons for the excess returns are set at one and three months. The sample period used to generate the forecasts covers three months, with the predictions implemented based on a rolling window procedure.

We use four indicators—buying-in-good-time, selling-in-bad-time, and right-decision probabilities and the return-to-volatility ratio—to determine which strategy is better. Buying-in-good-time (Selling-in-bad-time) probability is defined as the likelihood of the next month S&P 500 index return is positive (negative) conditional on the event that we decide to buy (sell). Right-decision probability is defined as the likelihood that the direction of our trade is consistent with that of the next month S&P 500 return conditional on the event that we decide to buy or sell. Returns-to-volatility ratio is based on the mean and standard deviation of the monthly returns created from the strategy.

Figures 1 and 2 compares the performance of models with $VIX^{2}_{t,30}$, $DVIX^{2}_{t,30}$, or both and show the results for the one- and three-month updating frequencies, respectively. Figure 1

14 We follow the estimation of Borkovec and Serbin (2013) to set the transaction cost as 10 basis points.
shows that for the one-month frequency the overall the model with $DVIX_{t,30}^2$ outperforms the model with $VIX_{t,30}^2$. The model with both information variables underperforms compared to the other two models. In particular, in terms of either the return-to-volatility ratio or the critical value from 0.01 to 0.05, the model with $DVIX_{t,30}^2$ performs better than the model with $VIX_{t,30}^2$, implying that the VIX term structure information provides a more precise prediction than the VIX level when the S&P 500 index drops. In the same vein, the performance of the model with $DVIX_{t,30}^2$ is superior when the other three measures are adopted respectively.

Figure 2 provides similar findings for the three-month frequency. The model with $DVIX_{t,30}^2$ dominates the model with $VIX_{t,30}^2$ in terms of the return-to-volatility ratio. Moreover, the buying-in-good-time of the model with $DVIX_{t,30}^2$ is higher than that with $VIX_{t,30}^2$.

In sum, a trading strategy that follows the forecasts generated from the information on the VIX term structure clearly outperforms a strategy that follows the forecasts generated from only the VIX level. Therefore, although the VIX level is also informative to some extent, both the in-sample regression analysis and the out-of-the-sample trading strategy highlight the merits of the information implied in the VIX term structure with regard to the prediction of
excess returns in the S&P 500 index.

4.6. Robustness Analysis

To ensure that we have a sufficiently large number of observations, we run the regression models using overlapping data (i.e., daily one-month returns). We adopt robust Newey–West standard errors for our analyses. To determine whether the findings are sensitive to data selection, we rerun the regressions using non-overlapping data. Table 7 reports the results of the full model for one-, two-, and three-month prediction horizons.

<TABLE 7 ABOUT HERE>

In general, Table 7 shows that the variable representing the information on the squared VIX term structure is more informative than the variable representing the squared VIX level information. The coefficient on the term structure variable, $\beta_2$, for the one- and three-month prediction horizons is always significantly positive, a factor that is not dependent on the approach adopted for the compilation of the information variables. The coefficient on the level variable, $\beta_1$, is positive in all cases but insignificant. Only one coefficient, $PC1$ and $PC2$ as the term structure variable, is insignificant for the two-month prediction.

Overall, our findings based non-overlapping data are, in general, consistent with our
analysis using overlapping data. Both analyses provide support for the important role of the squared VIX term structure in the determination of future excess returns in the S&P 500 index. The information content of the squared VIX term structure is clearly significant and strongly complementary to that of the squared VIX level.

5. CONCLUSIONS

We construct a theoretical model based on Heston’s (1993) SV model to investigate the relation of the squared VIX level and VIX term structure with S&P 500 index excess returns. We find a positive relation between excess returns and the squared VIX level, which is consistent with the traditional capital asset pricing model. We also report a positive relation between excess market returns and the squared VIX term structure, which is consistent with the empirical results of Bakshi et al. (2011).

We use three alternative empirical approaches to support our theoretical model and find that the squared VIX term structure is more informative than the squared VIX level, although both factors have predictive power across various time horizons. The incremental contribution of the VIX term structure with regard to the prediction of excess returns is particularly discernible for shorter horizons. Finally, our results show that the adoption of trading strategies based on the forecasts generated from the information on the VIX term structure are clearly superior to those based on only the VIX level.
REFERENCES

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), ‘The Distribution of Realized

Studies, 20: 651-707.

Bakshi, G.S., N. Kapadia and D. Madan (2003), ‘Stock Return Characteristics, Skew Laws and
the Differential Pricing of Individual Equity Options’, Review of Financial Studies, 16:
101-43.

Bakshi, G.S., G. Panayotov and G. Skoulakis (2011), ‘Improving the Predictability of Real
Economic Activity and Asset Returns with Forward Variances inferred from Option


Information Content of Implied Volatilities and High-frequency Index Returns’, Journal

Bollerslev, T., G. Tauchen and H. Zhou (2009), ‘Expected Stock Returns and Variance Risk

Borkvec, M. and V. Serbin (2013), ‘Create or Buy: A Comparative Analysis of Liquidity and


Appendix

We follow the CBOE methodology to calculate the squared VIX as

$$VIX^2_{t, \tau} \equiv 365 \times \left( \frac{2}{\tau} \sum_{i} \frac{\Delta K_i}{K_i^2} e^{r \cdot \tau} \frac{\tau}{365} Q(K_i) - \frac{1}{\tau} \left( \frac{F}{K_0} - 1 \right)^2 \right).$$

(A.1)

where $K_i$ is the strike price of the $i$th out-of-the-money option; $\Delta K_i$ is the interval between two strike prices, defined as $\Delta K_i = (K_{i.1} - K_{i.1})/2$; in particular, $\Delta K_i$ for the lowest strike price is simply the difference between the lowest and the next higher strike price, that is, $K_{i.1} - K_i$. Similarly, $\Delta K_i$ for the highest strike price is $K_i - K_{i.1}$. $r_f$ refers to the risk-free rate; $\tau$ is the time to expiration defined as the number of calendar days; $Q(K_i)$ is the midpoint of the bid–ask spread for each option with strike price $K_i$; $F$ refers to the implied forward index level derived from the nearest-the-money index option prices based upon put–call parity; and $K_0$ is the first strike price below the forward index level.

We use the interpolation similar to that suggested by the CBOE to construct the VIX term structure with six maturities as

$$VIX^2_{t, \tau} = \left[ T_1 \times VIX^2_{t, T_1} \left( \frac{T_2 - \tau}{T_2 - T_1} \right) + T_2 \times VIX^2_{t, T_2} \left( \frac{T_2 - \tau}{T_2 - T_1} \right) \right] \times \frac{1}{\tau}$$

(A.2)

where $\tau = 30, 60, 90, 180, 270, \text{ or } 360$ days, and $T_1$ and $T_2$ are the two nearest maturities embracing $\tau$. We use a similar method to calculate the term structure of the forward squared VIX as

$$FVIX^2_{t, T_1, T_2} = VIX^2_{t, T_1} \left( \frac{T_2}{T_2 - T_1} \right) - VIX^2_{t, T_1} \left( \frac{T_1}{T_2 - T_1} \right), T_1 < T_2.$$
Table 1. Correlation matrix of excess future monthly returns, VIX30 and DVIX30

$ER_{t,t+30}$ is the excess return for the period from time $t$ to $t+30$, defined as the return on the S&P 500 index minus the three-month T-bill rate; $VIX_{t,30}^2$ is the squared VIX with a 30 calendar-day maturity; and $DVIX_{t,30}^2$ is the difference between the squared VIX with a 60 calendar-day maturity and the squared VIX with a 30 calendar-day maturity, scaled by 2; that is, $DVIX_{t,30}^2 = 2(VIX_{t,60}^2 - VIX_{t,30}^2)$.

<table>
<thead>
<tr>
<th></th>
<th>$ER_{t,t+30}$</th>
<th>$VIX_{t,30}^2$</th>
<th>$DVIX_{t,30}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ER_{t,t+30}$</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_{t,30}^2$</td>
<td>0.0047</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$DVIX_{t,30}^2$</td>
<td>0.1132</td>
<td>-0.6167</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 2  Correlation matrix of squared VIX values under different horizons

$VIX_t^2,h$ is the squared VIX with an $h$ calendar-day maturity, where $h = 30, 60, 90, 180, 270, \text{ and } 360$.

<table>
<thead>
<tr>
<th></th>
<th>$VIX_{t,30}^2$</th>
<th>$VIX_{t,60}^2$</th>
<th>$VIX_{t,90}^2$</th>
<th>$VIX_{t,180}^2$</th>
<th>$VIX_{t,270}^2$</th>
<th>$VIX_{t,360}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VIX_{t,30}^2$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_{t,60}^2$</td>
<td>0.9852</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_{t,90}^2$</td>
<td>0.9622</td>
<td>0.9899</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_{t,180}^2$</td>
<td>0.9031</td>
<td>0.9522</td>
<td>0.9767</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_{t,270}^2$</td>
<td>0.8519</td>
<td>0.9128</td>
<td>0.9460</td>
<td>0.9834</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$VIX_{t,360}^2$</td>
<td>0.7541</td>
<td>0.8222</td>
<td>0.8583</td>
<td>0.9108</td>
<td>0.9362</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 3. Forward implied variance predictions

This table presents the results based on the following regression model:

\[ ER_{t,30} = \alpha + \beta_1 VIX_{t,30}^2 + \beta_2 DVIX_{t,30}^2 + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon, \]

where \( ER_{t,30} \) is the excess return for the period from time \( t \) to \( t+30 \), defined as the return on the S&P 500 index minus the three-month T-bill rate; \( VIX_{t,30}^2 \) is the squared VIX with a 30 calendar-day maturity; and \( DVIX_{t,30}^2 \) is the difference between the squared VIX with a 60 calendar-day maturity and the squared VIX with a 30 calendar-day maturity, scaled by 2, that is: \( DVIX_{t,30}^2 = 2(\sqrt{\text{VIX}}_{t,60}^2 - \sqrt{\text{VIX}}_{t,30}^2) \), however, to avoid the collinearity, we use the residual of \( DVIX_{t,30}^2 \) regressed on \( VIX_{t,30}^2 \) to proxy for \( DVIX_{t,30}^2 \); \( VRP_t \) is the variance risk premium, that is: \( VRP_t = \text{VIX}_{t,30}^2 / 12 - \text{RV}_{t,30}^2 \), where \( \text{RV}_{t,30} \) is the realized variance for the period from \( t-30 \) to \( t \); \( SKEW_t \) (\( KURT_t \)) is the model-free skewness (kurtosis) with a 30 calendar-day maturity (Bakshi et al., 2003); \( TMS_t \) is the term spread defined as the difference between the ten-year and three-month Treasury yields; and \( DFS_t \) is the term spread defined as the difference between Moody’s BAA and AAA bond yield indices (Bollerslev et al., 2009). The standard errors are calculated based upon the Newey–West method, with the lag being equal to the number of overlapping horizons. The sample period runs from 2 January 1998 to 31 August 2012. ** and * indicate significance at the 5% and 10% levels, respectively.

<table>
<thead>
<tr>
<th>One-month horizon</th>
<th>Models</th>
<th>Coeff. (1)</th>
<th>S.E.</th>
<th>Coeff. (2)</th>
<th>S.E.</th>
<th>Coeff. (3)</th>
<th>S.E.</th>
<th>Coeff. (4)</th>
<th>S.E.</th>
<th>Coeff. (5)</th>
<th>S.E.</th>
<th>Coeff. (6)</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td></td>
<td>0.0022</td>
<td>0.0042</td>
<td>0.0024</td>
<td>0.0032</td>
<td>-0.0054</td>
<td>0.0154</td>
<td>-0.0054</td>
<td>0.0154</td>
<td>0.0060</td>
<td>0.0146</td>
<td>0.0102</td>
<td>0.0159</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td></td>
<td>0.0032</td>
<td>0.0879</td>
<td>-</td>
<td>-</td>
<td>0.0032</td>
<td>0.0736</td>
<td>0.0121</td>
<td>0.1052</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0831*</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td></td>
<td>-</td>
<td>-</td>
<td>0.4782*</td>
<td>0.2495</td>
<td>0.4782*</td>
<td>0.2505</td>
<td>-</td>
<td>-</td>
<td>0.5520**</td>
<td>0.2488</td>
<td>0.6195**</td>
<td>0.2583</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0266**</td>
<td>0.0133</td>
<td>0.0249*</td>
<td>0.0127</td>
<td>0.0268*</td>
<td>0.0136</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0034</td>
<td>0.0082</td>
<td>-0.0001</td>
<td>0.0070</td>
<td>0.0020</td>
<td>0.0070</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0001</td>
<td>0.0007</td>
<td>0.0003</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0007</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_6 )</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0012</td>
<td>0.0023</td>
<td>-0.0016</td>
<td>0.0024</td>
<td>-0.0014</td>
<td>0.0024</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_7 )</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0008</td>
<td>0.0139</td>
<td>-0.0075</td>
<td>0.0115</td>
<td>-0.0146</td>
<td>0.0151</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>\text{R}^2</td>
<td></td>
<td>0.0002</td>
<td>0.0217</td>
<td>0.0220</td>
<td>0.0129</td>
<td>0.0386</td>
<td>0.0413</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4. First and second principal component predictions

This table presents the results based on the following regression model:

\[ ER_{t,30} = \alpha + \beta_1 PC_1 + \beta_2 PC_2 + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon, \]

where \( ER_{t,30} \) is the excess return for the period from time \( t \) to \( t+30 \), defined as the return on the S&P 500 index minus the three-month T-bill rate; \( PC_1 \) and \( PC_2 \) are the first and second principal components; \( VRP_t \) is the variance risk premium, that is: \( VRP_t = \frac{VIX_t^{30}}{12} - RV_{t,30} \), where \( RV_{t,30} \) is the realized variance for the period from \( t-30 \) to \( t \); \( SKEW_t \) (KURT \_t) is the model-free skewness (kurtosis) with a 30 calendar-day maturity (Bakshi et al., 2003); \( TMS_t \) is the term spread defined as the difference between the ten-year and three-month Treasury yields; and \( DFS_t \) is the term spread defined as the difference between Moody’s BAA and AAA bond yield indices (Bollerslev et al., 2009). The standard errors are calculated based upon the Newey–West method, with the lag being equal to the number of overlapping horizons. The sample period runs from 2 January 1998 to 31 August 2012. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th>One-month horizon</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.0024</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.0012</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_6 )</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_7 )</td>
<td>–</td>
</tr>
<tr>
<td>No. of obs.</td>
<td>3,685</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.0028</td>
</tr>
</tbody>
</table>
Table 5. Two-factor stochastic volatility framework proxy predictions [record 20160606]

This table presents the results based upon the following regression model:

\[ \text{ER}_{t,30} = \alpha + \beta_1 V_t + \beta_2 (m_t - V_t) + \beta_3 \text{VRP}_t + \beta_4 \text{SKEW}_t + \beta_5 \text{KURT}_t + \beta_6 \text{TMS}_t + \beta_7 \text{DFS}_t + \varepsilon, \]

where \( \text{ER}_{t,30} \) is the excess return for the period from time \( t \) to \( t+30 \), defined as the return on the S&P 500 index minus the three-month T-bill rate; \( V_t \) and \( m_t \) are the respective instantaneous variance and stochastic central tendency; \( \text{VRP}_t \) is the variance risk premium, that is: \( \text{VRP}_t = \text{VIX}_t \cdot m_t / 12 - \text{RV}_{t,30} \), where \( \text{RV}_{t,30} \) is the realized variance for the period from \( t+30 \) to \( t \); \( \text{SKEW}_t \) (\( \text{KURT}_t \)) is the model-free skewness (kurtosis) with a 30 calendar-day maturity (Bakshi et al., 2003); \( \text{TMS}_t \) is the term spread defined as the difference between the ten-year and three-month Treasury yields; and \( \text{DFS}_t \) is the term spread defined as the difference between Moody’s BAA and AAA bond yield indices (Bollerslev et al., 2009). The standard errors are calculated based upon the Newey-West method, with the lag being equal to the number of overlapping horizons. The sample period runs from 2 January 1998 to 31 August 2012. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.0030</td>
<td>0.0037</td>
<td>0.0024</td>
<td>0.0032</td>
<td>-0.0086</td>
<td>0.0066</td>
<td>0.0002</td>
<td>0.0155</td>
<td>0.0010</td>
<td>0.0116</td>
<td>0.0194</td>
<td>0.0122</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.0141</td>
<td>0.0760</td>
<td>-</td>
<td>-</td>
<td>0.1780</td>
<td>0.1140</td>
<td>0.0314</td>
<td>0.0888</td>
<td>-</td>
<td>-</td>
<td>0.4181***</td>
<td>0.1612</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-</td>
<td>-</td>
<td>0.0981</td>
<td>0.0791</td>
<td>0.2730**</td>
<td>0.1330</td>
<td>-</td>
<td>-</td>
<td>0.1800***</td>
<td>0.0970</td>
<td>0.4033***</td>
<td>0.1475</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0289**</td>
<td>0.0134</td>
<td>0.0200*</td>
<td>0.0120</td>
<td>0.0148*</td>
<td>0.0082</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0016</td>
<td>0.0023</td>
<td>-0.0010</td>
<td>0.0021</td>
<td>-0.0001</td>
<td>0.0019</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0002</td>
<td>0.0004</td>
<td>-0.0002</td>
<td>0.0004</td>
<td>-0.0005</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \beta_6 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0008</td>
<td>0.0023</td>
<td>-0.0011</td>
<td>0.0023</td>
<td>-0.0016</td>
<td>0.0022</td>
</tr>
<tr>
<td>( \beta_7 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0001</td>
<td>0.0133</td>
<td>-0.0044</td>
<td>0.0132</td>
<td>-0.0184</td>
<td>0.0142</td>
</tr>
</tbody>
</table>

| \( R^2 \)         | 0.0001     | 0.0108| 0.0241     | 0.0142| 0.0259     | 0.0448| 0.0448     | 0.0448| 0.0448     | 0.0448| 0.0448     | 0.0448|
Table 6. Time horizon predictions

This table presents the results based on the following regression models:

\[
ER_{t,h} = \alpha + \beta_1 VIX_{t,30} + \beta_2 DVIX_{t,30} + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon \quad \text{(Model 1)}
\]

\[
ER_{t,h} = \alpha + \beta_1 PC1_t + \beta_2 PC2_t + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon \quad \text{(Model 2)}
\]

\[
ER_{t,h} = \alpha + \beta_1 V_t + \beta_2 (m_t - V_t) + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon \quad \text{(Model 3)}
\]

where \( h = 30, 60, 90, 180, 270, \) and 360, with the sample period running from 2 January 1998 to 31 August 2012. The table presents the coefficients on only the main explanatory variables, \( VIX_{t,30}, DVIX_{t,30}, PC1_t, PC2_t, V_t \) and \( m_t - V_t \), along with the standard errors, which are calculated based upon the Newey–West method, where the lags are equal to the number of overlapping horizons. We present the \( R^2 \) of the regression models with the two main factors and all of the control variables. Each panel presents the \( R^2 \) of the regression models with only the level factor and all control variables and the ratio of \( R^2 \) to \( R^2 \) with the level factor and all control variables minus 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>One-month</th>
<th>Two-month</th>
<th>Three-month</th>
<th>Six-month</th>
<th>Nine-month</th>
<th>One-year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
<td>S.E.</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.0831*</td>
<td>0.0437</td>
<td>0.2334*</td>
<td>0.1019</td>
<td>0.2947*</td>
<td>0.1567</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>0.6195**</td>
<td>0.2583</td>
<td>0.8345**</td>
<td>0.3873</td>
<td>1.1422***</td>
<td>0.4690</td>
</tr>
<tr>
<td>No. of Obs.</td>
<td>3,685</td>
<td>3,643</td>
<td>3,622</td>
<td>3,559</td>
<td>3,496</td>
<td>3,441</td>
</tr>
<tr>
<td>( R^2 ) with term structure &amp; controls</td>
<td>0.0413</td>
<td>0.0615</td>
<td>0.1095</td>
<td>0.1703</td>
<td>0.1961</td>
<td>0.2233</td>
</tr>
<tr>
<td>( R^2 ) with VIX_{t,30} &amp; controls</td>
<td>0.0129</td>
<td>0.0353</td>
<td>0.0758</td>
<td>0.1071</td>
<td>0.1421</td>
<td>0.1874</td>
</tr>
<tr>
<td>( R^2 ) ratio -1</td>
<td>2.20</td>
<td>0.74</td>
<td>0.44</td>
<td>0.59</td>
<td>0.38</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 6 continues
Table 6 (cont.)

<table>
<thead>
<tr>
<th>Variables</th>
<th>One-month</th>
<th>Two-month</th>
<th>Three-month</th>
<th>Six-month</th>
<th>Nine-month</th>
<th>One-year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
<td>S.E.</td>
</tr>
<tr>
<td>Panel B: Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0031**</td>
<td>0.0016</td>
<td>0.0084**</td>
<td>0.0041</td>
<td>0.0120</td>
<td>0.0066</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0130**</td>
<td>0.0062</td>
<td>0.0153**</td>
<td>0.0087</td>
<td>0.0212**</td>
<td>0.0102</td>
</tr>
<tr>
<td>No. of Obs.</td>
<td>3.685</td>
<td>3.643</td>
<td>3.622</td>
<td>3.559</td>
<td>3.496</td>
<td>3.441</td>
</tr>
<tr>
<td>$R^2$ with term structure &amp; controls</td>
<td>0.0368</td>
<td>0.0639</td>
<td>0.1130</td>
<td>0.1906</td>
<td>0.2145</td>
<td>0.2372</td>
</tr>
<tr>
<td>$R^2$ with $VIX_{t-30}^2$ &amp; controls</td>
<td>0.0188</td>
<td>0.0513</td>
<td>0.0882</td>
<td>0.1522</td>
<td>0.1867</td>
<td>0.2179</td>
</tr>
<tr>
<td>$R^2$ ratio -1</td>
<td>0.96</td>
<td>0.25</td>
<td>0.28</td>
<td>0.25</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>Panel C: Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.4181***</td>
<td>0.1612</td>
<td>0.8030***</td>
<td>0.2490</td>
<td>1.2630***</td>
<td>0.3690</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.4033***</td>
<td>0.1475</td>
<td>0.7280***</td>
<td>0.2310</td>
<td>1.0280***</td>
<td>0.2990</td>
</tr>
<tr>
<td>No. of Obs.</td>
<td>3.686</td>
<td>3.643</td>
<td>3.622</td>
<td>3.559</td>
<td>3.496</td>
<td>3.441</td>
</tr>
<tr>
<td>$R^2$ with term structure &amp; controls</td>
<td>0.0448</td>
<td>0.0764</td>
<td>0.1287</td>
<td>0.2056</td>
<td>0.2231</td>
<td>0.2374</td>
</tr>
<tr>
<td>$R^2$ with $VIX_{t-30}^2$ &amp; controls</td>
<td>0.0142</td>
<td>0.0332</td>
<td>0.0766</td>
<td>0.1006</td>
<td>0.1343</td>
<td>0.1822</td>
</tr>
<tr>
<td>$R^2$ ratio -1</td>
<td>2.15</td>
<td>1.30</td>
<td>1.67</td>
<td>1.04</td>
<td>0.66</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Table 7. Predictions with non-overlapping data

This table presents the results based upon the following regression models:

$$ER_{t,h} = \alpha + \hat{\beta}_1VIX_{t-30} + \hat{\beta}_2DVIX_{t-30} + \hat{\beta}_3VRP_t + \hat{\beta}_4SKEW_t + \hat{\beta}_5KURT_t + \hat{\beta}_6TMS_t + \hat{\beta}_7DFS_t + \epsilon_t \quad \text{(Model 1)}$$

$$ER_{t,h} = \alpha + \hat{\beta}_1PC1_t + \hat{\beta}_2PC2_t + \hat{\beta}_3VRP_t + \hat{\beta}_4SKEW_t + \hat{\beta}_5KURT_t + \hat{\beta}_6TMS_t + \hat{\beta}_7DFS_t + \epsilon_t \quad \text{(Model 2)}$$

$$ER_{t,h} = \alpha + \hat{\beta}_1V_t + \hat{\beta}_2(m_t - V_t) + \hat{\beta}_3VRP_t + \hat{\beta}_4SKEW_t + \hat{\beta}_5KURT_t + \hat{\beta}_6TMS_t + \hat{\beta}_7DFS_t + \epsilon_t \quad \text{(Model 3)}$$

where $h = 30, 60, \text{ and } 90$, and the sample is non-overlapping, with the sample period running from 2 January 1998 to 31 August 2012. Standard errors are obtained based on the ordinary least square method.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Horizons</th>
<th>One-month</th>
<th>Two-month</th>
<th>Three-month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
<td>S.E.</td>
</tr>
<tr>
<td>Panel A: Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0189</td>
<td>0.0194</td>
<td>0.0556</td>
<td>0.0234</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.1198</td>
<td>0.1161</td>
<td>0.1860</td>
<td>0.2868</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.7711***</td>
<td>0.2774</td>
<td>2.5817***</td>
<td>0.7718</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.0123*</td>
<td>0.0072</td>
<td>0.0099*</td>
<td>0.0054</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.0087</td>
<td>0.0209</td>
<td>-0.0025</td>
<td>0.0412</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.0019</td>
<td>0.0029</td>
<td>0.0013</td>
<td>0.0047</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.0015</td>
<td>0.0035</td>
<td>-0.0072</td>
<td>0.0066</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.0197</td>
<td>0.0141</td>
<td>-0.0358</td>
<td>0.0253</td>
</tr>
<tr>
<td>No. of obs.</td>
<td>176</td>
<td>88</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0533</td>
<td>0.1358</td>
<td>0.2322</td>
<td></td>
</tr>
<tr>
<td>Panel B: Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0172</td>
<td>0.0234</td>
<td>0.0434</td>
<td>0.0491</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0010</td>
<td>0.0037</td>
<td>0.0098</td>
<td>0.0065</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0180**</td>
<td>0.0073</td>
<td>0.0269</td>
<td>0.0198</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.0152</td>
<td>0.0103</td>
<td>0.0053</td>
<td>0.0046</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.0051</td>
<td>0.0213</td>
<td>-0.0012</td>
<td>0.0412</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.0017</td>
<td>0.0030</td>
<td>0.0018</td>
<td>0.0051</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.0020</td>
<td>0.0036</td>
<td>-0.0072</td>
<td>0.0072</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.0140</td>
<td>0.0144</td>
<td>-0.0442</td>
<td>0.0286</td>
</tr>
<tr>
<td>No. of obs.</td>
<td>176</td>
<td>88</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0231</td>
<td>0.0884</td>
<td>0.1126</td>
<td></td>
</tr>
<tr>
<td>Panel C: Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0038</td>
<td>0.0123</td>
<td>0.0021</td>
<td>0.0257</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.1840</td>
<td>0.2030</td>
<td>0.1027</td>
<td>0.4460</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.3610**</td>
<td>0.1530</td>
<td>1.0845***</td>
<td>0.3851</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.0081*</td>
<td>0.0047</td>
<td>0.0025*</td>
<td>0.0014</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.0099</td>
<td>0.0213</td>
<td>0.0042</td>
<td>0.0406</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.0019</td>
<td>0.0029</td>
<td>0.0016</td>
<td>0.0050</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.0019</td>
<td>0.0035</td>
<td>-0.0081</td>
<td>0.0070</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.0157</td>
<td>0.0142</td>
<td>-0.0426</td>
<td>0.0280</td>
</tr>
<tr>
<td>No. of obs.</td>
<td>176</td>
<td>88</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0631</td>
<td>0.0826</td>
<td>0.1263</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Results of trading strategies (one-month)

Note: The strategy, based upon turnover rates of 30 calendar days and an estimation period of 90 calendar days, is to take up long (short) positions in the S&P 500 index when estimated returns exceed (are below) the critical value $x$ ($-x$), and do nothing when estimated returns are in the interval, $[-x, x]$. The transaction cost is set to be 10 bps for each transaction. The strategy is based on the following estimated regression model:

$$ ER_{s,sh} = \alpha + \beta_1 VIX_{t,30}^2 + \beta_2 DVIX_{t,30}^2 + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon_t. $$
Figure 2. Results of trading strategies (three-month)

Note: The strategy, based upon turnover rates of 90 calendar days and an estimation period of 90 calendar days, is to take up long (short) positions in the S&P 500 index when estimated returns exceed (are below) the critical value \( x \) (\( -x \)), and do nothing when estimated returns are in the interval, \([-x, x]\). The transaction cost is set to be 10 bps for each transaction. The strategy is based upon the following estimated regression model:

\[
ER_{t,3m} = \alpha + \beta_1 VIX_{t,30}^2 + \beta_2 DVIX_{t,30}^2 + \beta_3 VRP_t + \beta_4 SKEW_t + \beta_5 KURT_t + \beta_6 TMS_t + \beta_7 DFS_t + \epsilon_t.
\]