Received: 19 June 2017 | Accepted: 27 November 2017
DOI: 10.1002/fut.21903

RESEARCH ARTICLE

WILEY FUTURES MARKETS

Are there gains from using information over the surface of implied volatilities?

Biao Guo[1] | **Qian Han**[2] | **Hai Lin**[3]

[1] School of Finance, Renmin University of China, Beijing, China

[2] Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China

[3] School of Economics and Finance, Victoria University of Wellington, Wellington, New Zealand

Correspondence
Hai Lin, School of Economics and Finance, Victoria University of Wellington, Wellington 6140, New Zealand.
Email: hai.lin@vuw.ac.nz

Funding information
National Natural Science Fund of China, Grant numbers: 71471153, 71503254; NNSFC Key Project, Grant number: 71631004

We investigate the out-of-sample predictability of implied volatility using the information over the implied volatility surface. We show that implied volatility surface is useful for the out-of-sample forecast of implied volatility up to 1 week ahead. Trading strategies based on the predictability of implied volatility could generate significant risk-adjusted gains after controlling for transaction costs. Significant results also depend on the way of modeling implied volatility surface. We then calibrate a two-factor stochastic volatility option pricing model to implied volatility data. Results show that implied volatility is better explained by both long- and short-term variance factors.

KEYWORDS
economic significance, implied volatility, out-of-sample forecast, two-factor stochastic volatility model

1 | INTRODUCTION

Whether asset returns are predictable has been a longstanding research question in literature.[1] On option market, Harvey and Whaley (1992), Gonclaves and Guidolin (2006), Konstantinidi, Skiadopoulos, and Tzagkaraki (2008), Chalamandaris and Tsekrekos (2010, 2011) and Neumann and Skiadopoulos (2013) find that option implied volatilities are statistically predictable. However, the economic profits become insignificant once the transaction costs are accounted for. Literature documents a disparity between statistical and economical significance of option market predictability.[2]

In this paper, we solve the disparity by using implied volatility surface information. The trading of the option market is dominated by short-maturity options. Nevertheless, Bakshi, Cao, and Chen (1997) find that long-dated options have information not readily available from short-dated options. Recently, Christoffersen, Jacobs, Ornthanalai, and Wang (2008) and Christoffersen, Heston, and Jacobs (2009) proposed component volatility models, and decomposed stochastic volatility into long- and short-term components. They find that component volatility models perform better than one-factor stochastic volatility

[1]See, for example, Fama and Schwert (1977), Fama and French (1988), Campbell and Shiller (1988), Kothari and Shanken (1997), Rapach et al. (2010, 2013), Pettenuzzo, Timmermann, and Valkanov (2014), Rapach, Ringgenberg, and Zhou (2016) on predicting stock returns; Keim and Stambaugh (1986), Fama and French (1989), Greenwood and Hanson (2013), Lin et al. (2014), Lin, Wu, and Zhou (2017) on predicting corporate bond returns; and Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005), Goh, Jiang, Tu, and Zhou, (2012), Sarno, Schneider and Wagner (2016), Gargano, Pettenuzzo, and Timmermann (2017), Lin, Liu, Wu, and Zhou (2017) on predicting Treasury bond returns.

[2]Similar disparity of statistical and economic significance on Treasury return predictability is documented in Thornton and Valente (2012).

model. These findings suggest there exists useful information in the whole implied volatility surface. Bakshi et al. (1997), Christoffersen et al. (2008, 2009) analyze the statistical significance. We extend their analysis to investigate the economic significance of implied volatility surface, and document significant economic gains by using the information of implied volatility surface.

We test whether incorporating the information of implied volatility surface can improve the prediction of implied volatility. If both long- and short-maturity implied volatilities contain useful information, using the whole implied volatility surface information will be able to improve the volatility forecast that is only based on one particular maturity information. We examine 14 models and compare their out-of-sample performance with that of the benchmark AR(1) model. These competing models are two adapted Nelson and Siegel models used by Diebold and Li (2006) for Treasury securities and by Chalamandaris and Tsekrekos (2011) for currency options, six time series models similar to Diebold and Li (2006), five combination models as in Rapach, Strauss, and Zhou (2010) and a Mallows model averaging (MMA) combination as in Hansen (2007, 2008). We use the implied volatility surface information of the at-the-money (ATM) options and the options with $\Delta 0.40$ and $\Delta 0.60$. We choose call option in our main analysis, and use put option as a robustness check. We find that, historical surface information plays a significant role in the prediction of implied volatilities. When daily data are used to forecast the 30-day implied volatility 1 day ahead and 5 days ahead, the best out-of-sample R^2 value is as high as 7.39% and 7.64%, respectively.[3] Results are significant across almost all maturities. Our results reveal the importance of using the whole implied volatility surface information. However, these models lose their predictive power beyond a week, suggesting that only the historical information within 1 week of the forecast date is important for the short-term forecast of index option market.

To examine whether the predictability has economic value, we construct a trading strategy based on a forecast by each model, and compare the portfolio performance with that of the benchmark AR(1) model. Using the gain on Leland's alpha (Leland, 1999) as the performance measure, we find that those models that utilize information from the entire surface generate significant economic profits up to 5 days ahead even after transaction costs are considered. For example, when daily data are used, the trading strategy based on the 1-day-ahead forecast by the VAR(1) model of volatility change (VARC) generates a gain on Leland's alpha of 11.13% relative to the benchmark, and is significant at the 1% level. The trading strategy based on the 5-day-ahead forecast by the VAR(1) model of volatility change (VARC) generates a gain on Leland's alpha of 2.13% relative to the benchmark, and is significant at the 10% level. Results are robust to the impact of transaction cost. This finding distinguishes our study from most other literature that finds no predictability of the option market after considering transaction costs.

Our findings are robust over time and over different options. A sub-sample analysis using data during the recent 2007–2009 financial crisis period finds that the predictability still exists during the crisis. Implied volatilities can still be predicted 5 days ahead. Moreover, their economic significance of 1-day-ahead forecast becomes stronger during the crisis. Analysis using put option data and data with a broader range of Δ further confirms our main results.

In order to explain why implied volatility surface information helps improve the forecast, we estimate a two-factor stochastic volatility option pricing model to extract a long-term and a short-term variance factor. Regressions of option implied volatilities on these two factors reveal that both variance factors are important to explain the time variations of implied volatility. Long-maturity implied volatilities are more associated with the long-term variance factor, while short-maturity implied volatilities are more related to the short-term variance factor. Both long- and short-maturity implied volatilities contain useful information of the implied volatility term structure. We are able to provide a better prediction by using them jointly.

Our study contributes to the literature in several ways. Our findings shed light on volatility modeling. We evaluate an extensive set of 14 models. Our finding that the whole implied volatility surface provides useful information in forecasting implied volatility suggests that a one-factor model is not sufficient for volatility modeling. In this regard, we provide empirical evidence consistent with the emerging component volatility models.

We document both statistical and economic significance of option market predictability using the information of implied volatility surface. This finding is different from literature that documents significant statistical predictability but fails to uncover the economic significance. This finding provides new insights to the economic profit by the predictability of implied volatility.[4]

Egloff, Leippold, and Wu (2010) and Johnson (2017) show that besides level, slope also helps predict the implied variance. We differ from them by considering more flexible models to use the information contained in the surface of implied volatilities. As a robustness check, we also compare the 14 models with the two-factor model that uses level and slope as the

[3]These results are higher than or comparable to studies on the predictability of other financial markets. See, for example, Gargano et al. (2017) on Treasury return predictability, Lin et al. (2014) and Lin, Wu, et al. (2017) on corporate bond return predictability, and Rapach et al. (2010), Pettenuzzo et al. (2014) and Rapach et al. (2016) on stock return predictability.

[4]Galai (1977), Chiras and Manaster (1978), Poon and Pope (2000) and Hogan, Jarrow, Teo, and Warachka, (2004) also find significant excess returns of option trading strategies even when transaction costs are considered.

predictors. Results continue to show that a more flexible model specification using VAR framework provides a superior forecasting power up to 1 week.

The rest of the paper is structured as follows. Section 2 introduces our empirical methodologies, including the 14 prediction models tested, the out-of-sample performance evaluation criteria, and the two-factor stochastic volatility option pricing model. Section 3 discusses the data and presents the empirical results of out-of-sample forecast. Section 4 provides several robustness checks, including a sub-sample analysis using data covering the recent crisis period, the out-of-sample performance of put options, the comparison with other benchmark, predictability using option data with a different Δ range and gain on alpha from a different asset pricing model. Section 5 reports the results of stochastic volatility model calibration. Section 6 concludes the paper.

2 | EMPIRICAL METHODOLOGY

In this section, we first explain the prediction models to be tested, and the statistical and economic significance measures for evaluating prediction performance. We then introduce the two-factor stochastic volatility option pricing model used to calibrate the term structure of implied volatilities.

2.1 | Out-of-sample forecast

We use out-of-sample forecast to test the importance of using the information of implied volatility surface. Suppose we have implied volatility data from time 1 to time T, and the out-of-sample forecast starts from time m. At any time t between m and T, we use the information up to time t to estimate the coefficients, and then use the estimated coefficients and information at time t to forecast the implied volatility h days ahead. At time $t+h$, we compare the forecast implied volatility and the realized implied volatility to calculate the out-of-sample forecast errors. This procedure is repeated from time m to $T-h$.

2.1.1 | Prediction models

The Nelson and Siegel (NS, 1987) model and its extension (Diebold & Li, 2006) are widely accepted by industry for forecasting the yield curve due to their simplicity and efficiency. The interest rate and implied volatility term structures are quite similar in many aspects (see Christoffersen et al., 2009; Derman, Kani, & Zou, 1996). Just as each Treasury security has a corresponding yield to maturity, each traded index option has a corresponding implied volatility. Both the yield curve and implied volatility term structure exhibit a high degree of time and cross-sectional variation. Since the NS model is an empirical model, it can be borrowed directly to model the term structure of implied volatilities. We fit the implied volatility curve with moneyness v, $\sigma_t^v(\tau)$, using the NS model,

$$\sigma_t^v(\tau) = \beta_{1t}^v + \beta_{2t}^v \frac{1-\exp^{(-\lambda_t\tau)}}{\lambda_t\tau} + \beta_{3t}^v\left(\frac{1-\exp^{(-\lambda_t\tau)}}{\lambda_t\tau} - \exp^{(-\lambda_t\tau)}\right), \tag{1}$$

where τ is time to maturity and parameters $\beta_{1t}^v, \beta_{2t}^v$, and β_{3t}^v are estimated by ordinary least squares (OLS) with λ_t fixed at a pre-specified value of 0.0147.[5] The loading on β_{1t}^v is 1, a constant that does not decay to 0 in the limit; hence β_{1t}^v may be viewed as a long-term factor. The loading on β_{2t}^v is $\frac{1-\exp^{(-\lambda_t\tau)}}{\lambda_t\tau}$, a function that starts at 1 but decays monotonically and quickly to 0; and hence may be viewed as a short-term factor. The loading on β_{3t}^v is $\frac{1-\exp^{(-\lambda_t\tau)}}{\lambda_t\tau} - \exp^{(-\lambda_t\tau)}$, which starts at 0 and increases, and then decays to 0, hence it may be viewed as a medium-term factor.[6]

Besides the NS model, we consider six time series models following Diebold and Li (2006), five combination models as in Rapach et al. (2010) and the MMA combination as in Hansen (2007, 2008). Table 1 lists all 14 models evaluated in this paper. By using dummy variables for the options of different moneyness, we are able to combine the volatility surface information that has advantage over the use of the volatility curve information alone. We use the the implied volatility surface information of the ATM options and the options with $\Delta 0.40$ and $\Delta 0.60$.[7] We define three dummy variables $ID_1 =$ if it is an ATM option and 0 otherwise, $ID_2 =$ if the option's $\Delta 0.40$ and 0 otherwise, $ID_3 =$ if the option's $\Delta 0.60$ and 0 otherwise. We use superscript v to

[5]Parameter λ_t governs the exponential decay rate; small values of λ_t produce slow decay and can better fit the curve for long maturities, while large values of λ_t produce fast decay and can better fit the curve for short maturities. Parameter λ_t also governs where the loading on β_{3t}^v achieves its maximum. As a result, we choose a λ_t value that maximizes the loading on the medium-term (122-day) factor, which gives 0.0147.

[6]Please refer to Guo, Han, and Zhou (2014) for a more detailed discussion.

TABLE 1 Prediction models

Model framework	Model No.	Model ID	Model description
Nelson-Siegel	(1)	NSAR	Nelson-Siegel factors as univariate AR(1) processes
Nelson-Siegel	(2)	NSVAR	Nelson-Siegel factors as multivariate VAR(1) processes
VAR	(3)	VARL	VAR(1) on volatility levels
VAR	(4)	VARC	VAR(1) on volatility changes
ECM	(5)	ECM1	ECM(1) with one common trend
ECM	(6)	ECM2	ECM(1) with two common trends
PCA	(7)	PCA	AR(1) regression on three principal components
Empirical component	(8)	EC	VAR(1) on empirical level, slope and curvature
Combination forecast	(9)	MC	Mean combination forecast
	(10)	MD	Median combination forecast
	(11)	TM	Trimmed mean combination forecast
	(12)	DMSPE1	DMSPE combination forecast with $\theta = 1$
	(13)	DMSPE2	DMSPE combination forecast with $\theta = 0.9$
	(14)	MMA	MMA combination forecast
Benchmark		AR(1) on volatility levels	

This table lists the 14 prediction models tested in this paper. The last row explains the benchmark model.

denote the moneyness. For implied volatility curve, we use the information of short-, medium- and long-maturity implied volatilities. In particular, we use 30-, 91-, 152-, 365-, and 730-day implied volatilities in the analysis. We then forecast the implied volatility surface h days ahead with the following models:

(1) Nelson-Siegel factors as univariate AR(1) processes (NSAR):

$$\hat{\sigma}^{v}_{t+h}(\tau) = \hat{\beta}^{v}_{1,\ t+h} + \hat{\beta}^{v}_{2,\ t+h} \frac{1-\exp^{(-\lambda_t \tau)}}{\lambda_t \tau} + \hat{\beta}^{v}_{3,\ t+h}\left(\frac{1-\exp^{(-\lambda_t \tau)}}{\lambda_t \tau} - \exp^{(-\lambda_t \tau)}\right), \tag{2}$$

where $\hat{\beta}^{v}_{i,t+h} = a_{1,i}ID_1 + a_{2,i}ID_2 + a_{3,i}ID_3 + \left(b_{1,i}ID_1 + b_{2,i}ID_2 + b_{3,i}ID_3\right)\hat{\beta}^{v}_{i,t}, i = 1, 2, 3.$ $a_{1i}a_{2i}a_{3i}$, $b_{1i}b_{2i}b_{3i}$ are all scalars.

(2) Nelson-Siegel factors as multivariate VAR(1) processes (NSVAR):

$$\hat{\sigma}^{v}_{t+h}(\tau) = \hat{\beta}^{v}_{1,t+h} + \hat{\beta}^{v}_{2,t+h} \frac{1-\exp^{(-\lambda_t \tau)}}{\lambda_t \tau} + \hat{\beta}^{v}_{3,t+h}\left(\frac{1-\exp^{(-\lambda_t \tau)}}{\lambda_t \tau} - \exp^{(-\lambda_t \tau)}\right), \tag{3}$$

where $\hat{\beta}^{v}_{t+h} = a_1 ID_1 + a_2 ID_2 + a_3 ID_3 + (b_1 ID_1 + b_2 ID_2 + b_3 ID_3)\hat{\beta}^{v}_{t}; \hat{\beta}^{v}_{t} = \begin{bmatrix} \hat{\beta}^{v}_{1t} & \hat{\beta}^{v}_{2t} & \hat{\beta}^{v}_{3t} \end{bmatrix} T$. a_1, a_2, a_3 are 3×1 vectors, and b_1, b_2, b_3 are 3×3 matrices.

(3) VAR(1) on volatility levels (VARL):

$\hat{\sigma}^{v}{}_{t+h} a_1 ID_1 + a_2 ID_2 + a_3 ID_3 + (b_1 ID_1 + b_2 ID_2 + b_3 ID_3)\sigma^{v}_{t}$, where

$\sigma^{v}{}_{t} = \begin{bmatrix} \sigma^{v}_{t}(30) \\ \sigma^{v}_{t}(91) \\ \sigma^{v}_{t}(152) \\ \sigma^{v}_{t}(365) \\ \sigma^{v}_{t}(730) \end{bmatrix}$. a_1, a_2, a_3 are 5×1 vectors, and b_1, b_2, b_3 are 5×5 matrices.

[7]This choice follows Bollen and Whaley (2004), Han (2007) and Yan (2011). For example, Bollen and Whaley (2004) and Han (2007) define ATM calls with Δ between 0.50 and 5/8 (approximately 0.60), and ATM puts with Δ between −0.50 and −3/8 (approximately −0.40). Yan (2011) defines OTM puts with Δ between −0.45 and −0.20. In the robustness check, we test the predictability using data with Δ between 0.30 and 0.70.

(4) VAR(1) on volatility changes (VARC): $\hat{z}^v{}_{t+h} = a_1ID_1 + a_2ID_2 + a_3ID_3 + (b_1ID_1 + b_2ID_2 + b_3ID_3)z^v{}_t$,

where $z^v{}_t = \begin{bmatrix} \sigma_t^v(30) - \sigma_{t-h}^v(30) \\ \sigma_t^v(91) - \sigma_{t-h}^v(91) \\ \sigma_t^v(152) - \sigma_{t-h}^v(152) \\ \sigma_t^v(365) - \sigma_{t-h}^v(365) \\ \sigma_t^v(730) - \sigma_{t-h}^v(730) \end{bmatrix}$. a_1, a_2, a_3 are 5×1 vectors, and b_1, b_2, b_3 are 5×5 matrices.

(5) ECM(1) with one common trend (ECM1): $\hat{z}^v{}_{t+h} = a_1ID_1 + a_2ID_2 + a_3ID_3 + (b_1ID_1 + b_2ID_2 + b_3ID_3)z^v{}_t$, where

$z^v{}_t = \begin{bmatrix} \sigma_t^v(30) - \sigma_{t-h}^v(30) \\ \sigma_t^v(91) - \sigma_t^v(30) \\ \sigma_t^v(152) - \sigma_t^v(30) \\ \sigma_t^v(365) - \sigma_t^v(30) \\ \sigma_t^v(730) - \sigma_t^v(30) \end{bmatrix}$. $a_1, a_2, a_3a_1, a_2, a_3$ are 5×1 vectors, and b_1, b_2, b_3 are 5×5 matrices.

(6) ECM(1) with two common trends (ECM2): $\hat{z}^v{}_{t+h} = a_1ID_1 + a_2ID_2 + a_3ID_3 + (b_1ID_1 + b_2ID_2 + b_3ID_3)z^v{}_t$,

where $z^v{}_t = \begin{bmatrix} \sigma_t^v(30) - \sigma_{t-h}^v(30) \\ \sigma_t^v(91) - \sigma_{t-h}^v(91) \\ \sigma_t^v(152) - \sigma_t^v(30) \\ \sigma_t^v(365) - \sigma_t^v(30) \\ \sigma_t^v(730) - \sigma_t^v(30) \end{bmatrix}$.$a_1, a_2, a_3$ are 5×1 vectors, and b_1, b_2, b_3 are 5×5 matrices.

(7) AR(1) regression on three principal components (PCA). We first conduct a principal component analysis on the volatility time series data. Denote the largest three eigenvalues by λ_1^v, λ_2^v, and λ_3^v, with associated eigenvectors q_1^v, q_2^v, and q_3^v, and the first three principal components $x^v{}_t = [x_{1t}^v \quad x_{2t}^v \quad {}^v x_{3t}]^{\mathrm{T}}$. We first forecast $x^v{}_{t+1}$ with an AR(1) model, $\hat{x}_{i,t+h}^v = a_{1,i}ID_1 + a_{2i}ID_2 + a_{3i}ID_3 + (b_iID_1 + b_iID_2 + b_iID_3)x_{i,t}^v, i = 1, 2, 3,$ and then generate forecasts for volatilities as $\hat{\sigma}_{t+h}^v(\tau) = q_1^v(\tau)\hat{x}_{1,t+h}^v + q_2^v(\tau)\hat{x}_{2,t+h}^v + q_3^v(\tau)\hat{x}_{3,t+h}^v$. $a_{1,i},\ a_{2,i},\ a_{3,i},\ b_{1,i},\ b_{2,i},\ b_{3,i}$ are all scalars.

(8) VAR(1) on empirical level, slope and curvature (EC): $\sigma_{t+h}^v = a_1ID_1 + a_2ID_2 + a_3ID_3 + (b_1ID_1 + b_2ID_2 + b_3ID_3)F^v{}_t$,

where $F^v{}_t = \begin{bmatrix} \sigma_t^v(365) \\ \sigma_t^v(365) - \sigma_t^v(30) \\ 2 \times \sigma_t^v(122) - \left(\sigma_t^v(365) + \sigma_t^v(30)\right) \end{bmatrix}$. We compute the empirical level, slope and curvature of the volatility term structure. The empirical level is defined as the 365-day implied volatility. The slope is the 365-day implied volatility minus the 30-day implied volatility. Finally, the curvature is two times the 122-day implied volatility minus the sum of the 365- and 30-day implied volatilities. a_1, a_2, a_3 are 3×1 vectors, and b_1, b_2, b_3 are 3×3 matrices.

Research has shown that combination forecasts typically outperform individual forecasts both statistically and economically. For example, Rapach et al. (2010) find that combinations deliver consistent forecast gains for equity premium predictions. Lin, Wang, and Wu (2014) document similar findings using corporate bond return data. So besides the forecasts in model (1) to (8), we further combine them as $\hat{\sigma}_{ct+h}^v(\tau) = \sum_{k=1}^{8} w_{k,t}^v(h,\tau)\hat{\sigma}_{k,t+h}^v(\tau)$, where $\hat{\sigma}_{k,t+h}^v(\tau)$ is the individual forecast using model k and $w_{k,t}^v(h,\tau)$ is the weight to be placed for the model-k forecast. Depending on the choice of weight $w_{k,t}^v(h,\tau)$, we provide five combination forecasts:

(9) The mean combination forecast (MC): $w_{k,t}^v(h,\tau) = 1/8$.

(10) The median combination forecast (MD): the median of $\hat{\sigma}_{k,t+h}^v(\tau), k = 1, 2, \ldots, 8$.

(11) The trimmed mean combination forecast (TM): $w_{k,t}^v(h,\tau) = 0$ for the smallest and largest forecasts and $w_{k,t}^v(h,\tau) = 1/6$ for the remaining forecasts.

(12) DMSPE (discount mean square prediction error) combination forecast one (DMSPE1): $w_{k,t}^v(h,\tau) = \frac{\left(\phi_{k,t}^v(h,\tau)\right)^{-1}}{\sum_{i=1}^{8}\left(\phi_{i,t}^v(h,\tau)\right)^{-1}}$,

where $\phi_{k,t}^v(h,\tau) = \sum_{j=m}^{t-h}\theta^{t-h-j}\left(\sigma_{j+h}^v(\tau) - \hat{\sigma}_{k,j+h}^v(\tau)\right)^2$. This weighting scheme gives more weight to the individual forecast with smaller out-of-sample prediction error. θ is a discounting factor deciding the size of the weights given to recent forecasts and m is the starting time of the out-of-sample forecast. We take $\theta = 1$ for no discounting to the remote forecast.

(13) DMSPE combination forecast two (DMSPE2): Same as model (12) except that $\theta = 0.9$ to give greater weight to recent forecasts.

Hansen (2007, 2008) proposes a forecast combination based on the MMA method. This method selects the forecast weight by minimizing a Mallow criterion that is a penalized sum of the square residuals. Hansen shows that MMA forecasts have better performance than other feasible forecasts.

(14) MMA combination (MMA). Let $w^v{}_t(h, \mathbf{t}) = \left[w^v_{1,t}(h,\tau)\ldots w^v_{8,t}(h,\tau)\right]T$ be the weight vector of the individual forecast, $\sigma^v{}_{j+h}(\tau) = \left[\hat{\sigma}^v_{1,j+h}(\tau)\ldots\hat{\sigma}^v_{8,j+h}(\tau)\right]^{\mathrm{T}}$ be the vector of the individual h-day ahead forecast of T-day implied volatility with moneyness v at time j, and $G = [g(1)\ldots g(8)]^{\mathrm{T}}$ be the vector of the predictor number used in the individual forecast. The MMA combination forecast set $w^v{}_t(h,\tau)$ to minimize $C^v_t(h,\tau)$ with the conditions that all $w^v_{k,t}(h,\tau)$ are non-negative and that $\sum_{k=1}^{8} w^v_{k,t}(h,\tau) = 1$. $C^v_t(h,\tau)$ is calculated by $C^v_t(h,\tau) = \sum_{j=m}^{t-h}\left(\sigma^v_{j+h}(\tau) - \sigma^v{}_{j+h}(\tau)^{\mathrm{T}} w^v_t(h,\tau)\right)^2 + 2w^v_t(h,\tau)^{\mathrm{T}} G s^2_v$, where s^2_v is an estimate of the variance of residuals from the largest fitted model.

The benchmark model is an AR(1) model on volatility levels: $\hat{\sigma}^v_{t+h}(\tau) = a(\tau) + b(\tau)\sigma^v_t(\tau)$. It only utilizes the time series information of an individual option series. If a model using the whole curve information significantly outperforms the benchmark, then we can be confident about the critical role of the surface information in implied volatility forecasting.

2.1.2 | Out-of-sample forecast evaluation

In order to check the performance of the prediction models relative to the benchmark model, we calculate the out-of-sample R^2 statistics of each model for each maturity across different moneyness, given by

$$R^2_{OS}(\tau) = 1 - \frac{\sum_v \sum_{j=m}^{T-h}\left(\sigma^v_{j+h}(\tau) - \hat{\sigma}^v_{j+h}(\tau)\right)^2}{\sum_v \sum_{j=m}^{T-h}\left(\sigma^v_{j+h}(\tau) - \bar{\sigma}^v_{j+h}(\tau)\right)^2}. \tag{4}$$

$\hat{\sigma}^v(\tau)$ and $\bar{\sigma}^v(\tau)$ are the forecast of implied volatility with moneyness v by model (1) to model (14) and the forecast by the benchmark AR(1) model, respectively. A positive $R^2_{OS}(\tau)$ value indicates that the prediction model outperforms the benchmark model. For model (12) and model (13) that require hold-out period (p) to calculate the optimal weight, the forecasting errors used to calculate the $R^2_{\mathrm{OS}}(\tau)$ values start from $m + p$ until $T - h$. We calculate the MSPE-adjusted statistic to test the significance of $R^2_{\mathrm{OS}}(\tau)$. Define

$$f_{t+h}(\tau) = \sum_v\left[\left(\sigma^v_{t+h}(\tau) - \bar{\sigma}^v_{t+h}(\tau)\right)\right]^2 - \sum_v\left[\left(\sigma^v_{t+h}(\tau) - \hat{\sigma}^v_{t+h}(\tau)\right)^2 - \left(\hat{\sigma}^v_{t+h}(\tau) - \bar{\sigma}^v_{t+h}(\tau)\right)^2\right], \tag{5}$$

and the MSPE-adjusted statistic is obtained by regressing $f_{t+h}(\tau)$ on a constant. The p-value corresponding to the constant from a one-sided test determines the significance of $R^2_{\mathrm{OS}}(\tau)$. We use Hodrick (1992) to calculate the standard errors that are robust to data overlap. To test the overall performance of the prediction models relative to the benchmark model, we also calculate the overall out-of-sample R^2 statistics of each model using

$$R^2_{\mathrm{OS}} = 1 - \frac{\sum_\tau \sum_v \sum_{j=m}^{T-h}\left(\sigma^v_{j+h}(\tau) - \hat{\sigma}^v_{j+h}(\tau)\right)^2}{\sum_\tau \sum_v \sum_{j=m}^{T-h}\left(\sigma^v_{j+h}(\tau) - \bar{\sigma}^v_{j+h}(\tau)\right)^2} \tag{6}$$

and test its significance with

$$f_{t+h} = \sum_\tau \sum_v\left[\left(\sigma^v_{t+h}(\tau) - \bar{\sigma}^v_{t+h}(\tau)\right)\right]^2 - \sum_\tau \sum_v\left[\left(\sigma^v_{t+h}(\tau) - \hat{\sigma}^v_{t+h}(\tau)\right)^2 - \hat{\sigma}^v_{t+h}(\tau) - \bar{\sigma}^v_{t+h}(\tau)^2\right]. \tag{7}$$

2.1.3 | Economic significance

We follow Cao and Han (2013) to evaluate the trading performance of the out-of-sample forecast and test whether the models generate abnormal profits. In order to avoid the potential problems by using the interpolated data in the economic significance analysis, we use S&P 500 index option transaction data to construct the portfolio. The trading strategies are simply based on the forecast volatility. Specifically, at date t we long (short) an option if the forecast volatility for that maturity at date $t+h$ is larger (smaller) than the current volatility. We delta-hedge our option position by buying (selling) Δ shares of S&P 500 index if we short (long) the options. The hedge ratio is calculated using the Black-Scholes option pricing formula. Its daily gain is calculated as

$$\pi_{i,t+1} = \left(\left(C_{i,t+1} - C_{i,t} \right) - \left((S_{t+1} - S_t) \Delta_{it} \right) - a \frac{r_{f,t}}{365} \left(C_{i,t} - \Delta_{it} S_t \right) \right. \tag{8}$$

C is the call option price, S is the S&P 500 index price, $r_{f,t}$ is the risk-free rate at date t, a is the number of calendar days between two trading dates. The same equation for the delta-hedged put options is applied, except we replace the call option price and delta with those of put option. We finally scale the dollar return $\pi_{i,t+1}$ by the absolute value of $C_{i,t} - \Delta_{it} S_t$ to convert to percentage return.

In order to compute returns of constant maturity, we construct option portfolios targeting maturities of 30, 91, 152, 365, and 730 days, with portfolio weights estimated in a way similar to Constantinides, Jackwerth, and Savov (2013). We form five portfolios made up of calls, each with targeted time to maturity of 30, 91, 152, 365, or 730 days. The weight of each option in a portfolio is calculated using an univariate Gaussian weighting kernel in days to maturity, with bandwidths of 10 days to maturity for the weighting kernel. We delete the options in a portfolio with weight smaller than 1% to remove outliers and normalize portfolio weights to sum to one. Thus, on any day our trading includes 15 portfolios with maturities of 30, 91, 152, 365, and 730 days, and moneyness of ATM, $\Delta 0.40$ and $\Delta 0.60$. We rebalance the portfolio daily and repeat the trade in the out-of-sample period. We then compound daily returns to h-day-ahead portfolio returns. We use the same approach to put options.

The gain on Leland's alpha is used to gauge the economic performance of these trading strategies. Leland's alpha takes into account the deviation of portfolio returns from normal distribution and is $\alpha_p = E\left[r_p\right] - \beta_p \left(E\left[r_m - r_f\right] \right) - r_f$, where r_m denotes the market return approximated by the S&P 500 index return, $\beta_p = \frac{\mathrm{cov}\left(r_p, -(1+r_m)^{-\gamma}\right)}{\mathrm{cov}\left(r_m, -(1+r_m)^{-\gamma}\right)}$ measures the systematic risk and $\gamma = \frac{\ln(E[1+r_m]) - \ln\left(1+r_f\right)}{\mathrm{var}(\ln(1+r_m))}$ measures the relative risk aversion. For comparison, these measures are then subtracted by the corresponding Leland's alpha of the benchmark AR(1) model.[8] A larger-than-zero gain on Leland's alpha thus indicates that the trading strategy generates an excess risk-adjusted return over the benchmark model. We annualize the gain on Leland's alpha and test its significance with Newey and West (1987) t-statistics adjusted for serial correlation.

Transaction cost is an important factor we need to control when we compare the performance of two trading strategies. A positive gain on Leland's alpha of one trading strategy might be due to its more aggressive trading. Thus its economic significance disappears once the transaction cost is accounted for. In order to examine whether our results are robust to the impact of transaction cost, we follow Cao, Han, Tong, and Zhan (2017) and introduce the transaction cost into the trading. We first use the mid price (MidP) that is the mid-point of bid and ask price. It does not assume any transaction cost. We then consider the effective option spread to be 10%, 25%, and 100% of the quoted spread.[9]

2.2 | A two-factor stochastic volatility option pricing model

In order to measure the impact of different volatility components on the implied volatility term structure, we estimate a two-factor stochastic volatility option pricing model, as in Christoffersen et al. (2009), where the variance of the risk-neutral ex-dividend stock return is determined by two factors,

[8]By saying gains from using information over surface of implied volatilities, we are more interested in comparing the models that use surface information with the benchmark model that does not use it. Thus all the statistical and economic significance measure we report in the paper are the comparison results to reflect our research focus.

[9]Studies of effective spread on equity options include Mayhew (2002), De Fontnouvelle, Fisher, and Harris (2003) and Muravyev and Pearson (2016). Mayhew (2002) and De Fontnouvelle et al. (2003) find that the ratio of effective spread to the quoted spread is less than 50% for equity options. Muravyev and Pearson (2016) show that for the average trade, effective spreads that take account of trade timing are one-third smaller than the traditionally used effective spreads.

$$dS_t = rS_t dt + \sqrt{V_{1t}} S_t dz_{1t} + \sqrt{V_{2t}} S_t dz_{2t}, \tag{9}$$

$$dV_{1t} = (a_1 - b_1 V_{1t})dt + \sigma_1 \sqrt{V_{1t}} dz_{3t},$$

$$dV_{2t} = (a_2 - b_2 V_{2t})dt + \sigma_2 \sqrt{V_{2t}} dz_{4t},$$

where z_{1t} and z_{2t} are uncorrelated, the correlation between z_{1t} and z_{3t} is ρ_1 and the correlation between z_{2t} and z_{4t} is ρ_2. We define the factor that is more persistent (b closer to zero) as the long-term variance factor, while the other is the short-term variance factor. As shown in Christoffersen et al. (2009), European options can be valued by a closed-form formula under this framework.

3 | DATA AND EMPIRICAL RESULTS

Our sample includes the implied volatilities of S&P 500 index options from 1996 to 2015. We use the volatility surfaces taken from the Ivy DB OptionMetrics database, with 10 different maturities (30, 60, 91, 122, 152, 182, 273, 365, 547, and 730 days) on each observation date. Since not all maturities are traded on each date, OptionMetrics interpolates the surface to obtain the missing data. Table 2 reports the mean, maximum, minimum, standard deviation, and autocorrelation of implied volatilities of ATM call options with different maturities.[10] The volatility curve is upward sloping, and long-maturity implied volatilities have smaller standard deviations than short-maturity implied volatilities. For example, the 730-day implied volatility has a mean of 20.17% and a standard deviation of 4.43%, while the 30-day implied volatility has a mean of 18.83% and a standard deviation of 7.51%. The different persistence across maturities suggests a necessity to model the long- and short-maturity implied volatilities separately.

Figure 1 plots the time series of the implied volatilities of ATM call options. It is clear that volatilities are time varying, with spikes occurring between 1998 and 1999, between 2002 and 2003 and between 2008 and 2009. They reflect the impact of the Asian crisis, the accounting scandal and the credit crisis, respectively. In the following empirical studies, we focus on the implied volatilities of five different maturities (30, 91, 152, 365, and 730 days) to reduce the dimensions in panel data model.

Table 3 reports the trading summary of options with different maturities. We report both the trading volume and the open interest. The option data with a negative bid-ask spread, a negative trading volume and open interest or a negative implied volatility are excluded. The trading volume and open interest of ATM (call and put) options, call options with $\Delta 0.60$ and call options with $\Delta 0.40$ are calculated from the options with moneyness between 45% and 55%, between 55% and 65% and between 35% and 45%, respectively.

TABLE 2 Summary statistics

Maturity (days)	Mean (%)	Std. dev. (%)	Min. (%)	Max. (%)	ρ (10)	ρ (30)	ρ (60)	ρ (180)
30	18.83	7.51	8.14	74.83	0.89	0.75	0.61	0.34
60	19.07	6.87	9.08	67.22	0.92	0.80	0.66	0.38
91	19.21	6.45	9.70	60.45	0.93	0.82	0.69	0.41
122	19.33	6.09	10.23	57.44	0.94	0.84	0.71	0.43
152	19.44	5.78	10.45	53.84	0.94	0.85	0.73	0.45
182	19.54	5.56	10.60	50.38	0.95	0.86	0.75	0.46
273	19.70	5.17	10.96	46.48	0.95	0.88	0.78	0.49
365	19.81	4.96	11.25	44.48	0.96	0.89	0.79	0.50
547	20.02	4.60	11.61	40.19	0.96	0.90	0.81	0.52
730	20.17	4.43	11.74	38.66	0.96	0.90	0.81	0.53

This table reports the summary statistics (mean, standard deviation, minimum, maximum, autocorrelation with lags of 10, 30, 60, and 180 days) of the implied volatilities of ATM call options. The sample period is from 1996 to 2015.

[10]The results of other moneyness are close to those of ATM options.

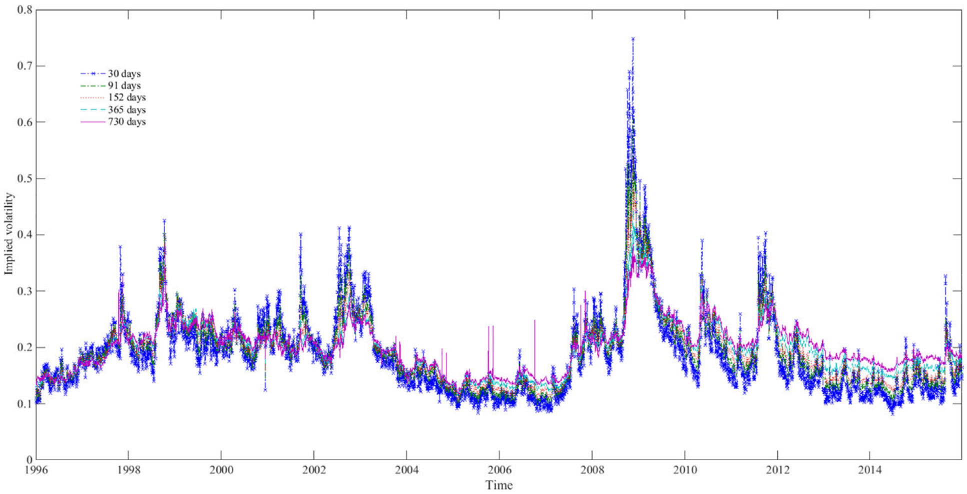


FIGURE 1 Implied volatility of selected maturities. This graph plots the time series of implied volatilities of selected maturities, specifically 30, 91, 152, 365, and 730 days. Sample period is from 1996 to 2015 [Color figure can be viewed at wileyonlinelibrary.com]

Trading in the option market is dominated by short-maturity options. For example, for the ATM call options, the options with maturities less than three months contribute about 79.39% to the total trading volume and about 54.90% to the total open interest. On the other hand, the options with maturities longer than 1 year only account for 2.96% of the total trading volume and 9.57% of the total open interest. The trade of long-maturity options is much less than that of short-maturity options. It is of great interest to investigate whether these limited trading contains useful information about future implied volatilities.

As a comparison, we also report the trading summary of call options with $\Delta 0.70$ using the options with moneyness between 65% and 75%, and $\Delta 0.30$ using the options with moneyness between 25% and 35%. Trading of call options with $\Delta 0.70$ is less active than those of $\Delta 0.60$, and dominated by short-maturity options. Nevertheless, trading of call options with $\Delta 0.30$ is more active than those of $\Delta 0.40$. This suggests that these options are liquid and frequently traded by investors.[11] In the robustness test, we add the options with $\Delta 0.70$ and $\Delta 0.30$ in the analysis to test whether the forecast result is robust to inclusion of more options.

We start the out-of-sample forecast in 2002. Parameters are estimated using a recursive window. Implied volatilities are forecast 1, 5, and 20 days ahead. The holdout out-of-sample period for model (12) and model (13) is set as 60 days.

We fit the implied volatility curve using the NS model by OLS on each observation date. Unreported results show that β_{1t}, as a long-term factor, displays a more persistent pattern than the other two factors. On the contrary, β_{2t} and β_{3t} are volatile since they represent the short and medium terms. They are especially pronounced when the market is turbulent. β_{1t} moves smoothly and captures the trend of the volatility very well, verifying that it reflects a long-term volatility. β_{2t} and VIX mimic each other, and taken together with the close movement between β_{2t} and the empirical slope lines, indicate that β_{2t} reflects the short-term volatility component and can be interpreted as a slope factor. The time variations of these factors provide some preliminary evidence that confirms the necessity of decomposing volatilities into long- and short-term components.[12]

3.1 | Statistical significance

Table 4 reports the out-of-sample R^2_{OS} statistics for all 14 models. The top, middle and bottom panels report the forecast results of 1, 5, and 20 days ahead, respectively. The results in the top panel show that most of our models beat the benchmark for a 1-day forecast. For example, 13 out of the 14 models generate a positive R^2_{OS} statistic at the 5% significance level or above for the 30-day implied volatility, and all combination forecasts have a greater than zero R^2_{OS} statistic and are significant at the 1% level. Similarly, there are 10 models that outperform the benchmark model at the 5% level or above for the 730-day implied volatility.

Among all the models, model VARC, which runs a VAR(1) model on the volatility change, and MMA combination, have the greatest R^2_{OS} value. Model ECM1 and model ECM2, which are ECM models with one and two common trends, respectively, also perform well. This suggests that the historical surface information is helpful when forecasting a particular maturity implied

[11]Thanks for the anonymous referee to point this out.

[12]In Appendix 1, we plot the time series of β_{1t}, β_{2t}, and β_{3t} of the NS model.

TABLE 3 Trading summary of S&P 500 index options

Maturity (days)	Volume	Percentage	Open interest	Percentage	Volume	Percentage	Open interest	Percentage
	ATM call				ATM put			
<30	40,826,704	29.12	193,057,678	20.17	42,851,640	30.27	180,550,841	21.07
30–91	70,480,199	50.27	332,449,354	34.73	72,196,523	51.00	326,092,925	38.06
91–152	14,868,885	10.61	123,813,077	12.93	15,524,699	10.97	122,601,772	14.31
152–365	9,870,019	7.04	216,370,952	22.60	8,436,133	5.96	169,304,814	19.76
365–730	3,257,799	2.32	77,318,610	8.08	2,169,028	1.53	52,129,581	6.08
>730	900,996	0.64	14,251,508	1.49	379,320	0.27	6,088,079	0.71
All	140,204,602	100.00	957,261,179	100.00	141,557,343	100.00	856,768,012	100.00
	Call with $\Delta = 0.60$				Call with $\Delta = 0.40$			
<30	23,023,845	38.11	195,034,840	24.31	34,707,239	44.34	189,011,008	22.96
30–91	25,997,656	43.03	308,898,741	38.50	29,377,294	37.53	288,242,921	35.02
91–152	4,682,105	7.75	99,185,971	12.36	5,848,476	7.47	113,299,573	13.76
152–365	4,439,338	7.35	147,334,272	18.36	6,264,739	8.00	178,055,275	21.63
365–730	1,800,099	2.98	45,091,625	5.62	1,778,007	2.27	48,277,885	5.86
>730	472,063	0.78	6,799,529	0.85	301,047	0.38	6,288,435	0.76
All	60,415,106	100.00	802,344,978	100.00	78,276,802	100.00	823,175,097	100.00
	Call with $\Delta = 0.70$				Call with $\Delta = 0.30$			
<30	13,014,136	55.83	201,197,166	33.64	37,280,577	46.16	203,664,354	24.08
30–91	7,809,005	33.50	244,844,064	40.94	28,084,411	34.78	285,230,619	33.72
91–152	980,076	4.20	65,617,431	10.97	5,995,761	7.42	119,034,913	14.07
152–365	1,081,000	4.64	70,052,792	11.71	7,302,547	9.04	188,065,864	22.23
365–730	354,853	1.52	14,594,019	2.44	1,799,856	2.23	43,772,154	5.17
>730	69,372	0.30	1,765,233	0.30	294,143	0.36	6,158,059	0.73
All	23,308,442	100.00	598,070,705	100.00	80,757,295	100.00	845,925,963	100.00

This table reports the trading volume and open interest of S&P 500 index options from 1996 to 2015. Option data with either a negative bid-ask spread, a negative trading volume and open interest or a negative implied volatility are excluded. The trading volume and open interest of ATM (call and put) options, call options with Δ equal to 0.60, 0.40, 0.70, and 0.30 are calculated from the options with moneyness between 45% and 55%, between 55% and 65%, between 35% and 45%, between 65% and 75% and between 25% and 35%, respectively.

volatility. It is interesting to observe that most of the R^2_{OS} statistics for the models using NS factors (model NSAR and model NSVAR) are negative, suggesting that they are not as good as the benchmark model in the out-of-sample period. This demonstrates the difference between in-sample fitting and out-of-sample forecast.

It is also interesting to observe that although forecast performance is improved overall by the use of the whole implied volatility surface information, not all models generate desirable results. Results using the principal components of the implied volatility curve (PCA and EC) are not significant. This finding is consistent with Kelly and Pruitt (2013, 2015) who claim that principal components may contain common error components that are irrelevant to forecasting, hence producing poor forecasting performance. Models PCA and EC use level information, while models VARC, ECM1, and ECM2 use the information of volatility changes that removes the volatility trend. Therefore, the performance difference across these models implies that both the information set and the way of modeling the information set are important when out-of-sample forecasts are performed on the option market.

Turning now to the performance of the 5-day-ahead forecast, the R^2_{OS} are smaller than those of 1-day-ahead forecast. There are nine models that have positive statistics that are significant at the 5% level for all maturities. The performance worsens for long-maturity implied volatilities. There is only one model (ECM1) that is significant at the 5% level for the 730-day implied volatility, while there are 10 such models for the 1-day-ahead forecast. However, models VARC, ECM1, and ECM2 continue to perform quite well for the 5-day-ahead forecast. The combination forecasts seem to deliver stable and significant results. Thus, in general, the implied volatility is still predictable 5 days ahead when we use daily data.

The bottom panel of Table 6 reports the results of the R^2_{OS} statistics of the 20-day-ahead forecast. It is clear that the forecasting abilities disappear and that none of the models is able to generate a positive R^2_{OS} statistic consistently across all maturities at the

TABLE 4 R^2_{OS} of implied volatility forecast

	Model													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Maturity days	**NSAR**	**NSVAR**	**VARL**	**VARC**	**ECM1**	**ECM2**	**PCA**	**EC**	**MC**	**MD**	**TM**	**DMSPE1**	**DMSPE2**	**MMA**
	One day ahead (%)													
30	−2.26	1.10^{a}	2.51^{a}	6.10^{a}	5.25^{a}	5.48^{a}	1.44^{a}	0.49^{a}	6.07^{a}	5.17^{a}	5.71^{a}	6.20^{a}	6.14^{a}	7.39^{a}
91	−12.59	−3.71	2.36^{a}	5.87^{a}	4.18^{a}	5.05^{a}	−7.15	−7.05	3.52^{a}	3.22^{a}	3.34^{a}	4.10^{a}	3.74^{a}	6.68^{a}
152	−9.14	−1.07	2.16^{a}	6.65^{a}	4.05^{a}	5.24^{a}	−8.38	−7.82	4.04^{a}	3.59^{a}	3.81^{a}	4.46^{a}	4.10^{a}	7.28
365	−25.35	−4.49	2.24^{a}	11.07^{a}	3.65^{a}	5.21^{a}	−9.26	−4.69	5.44^{a}	4.28^{a}	4.67^{a}	6.06^{a}	5.61^{a}	10.92^{a}
730	−5.96	−8.36	1.86^{a}	15.20^{a}	2.44^{a}	4.26^{a}	−9.76	−30.14	4.36^{a}	3.68^{a}	3.79^{a}	5.37^{a}	7.18^{a}	14.88^{a}
All	−7.54	−1.39	2.36^{a}	7.23^{a}	4.52^{a}	5.24^{a}	−3.42	−4.98	5.08^{a}	4.36^{a}	4.73^{a}	5.45^{a}	5.40^{a}	8.09^{a}
	Five days ahead (%)													
30	−0.57	1.85^{a}	0.84^{a}	6.33^{a}	4.72^{a}	4.83^{a}	2.86^{a}	0.83^{b}	5.57^{a}	4.13^{a}	4.92^{a}	5.68^{a}	6.54^{a}	7.64^{a}
91	−9.94	−3.56	−1.55	4.53^{a}	2.1^{a}	2.18^{a}	−2.89	−3.44	1.90^{a}	0.73^{b}	1.52^{a}	2.10^{a}	3.24^{a}	4.87^{a}
152	−7.57	−0.38	−1.58	4.27^{a}	1.77^{a}	1.76^{a}	−1.54	−2.34	2.56^{a}	1.54^{b}	2.17^{a}	2.66^{a}	3.67^{a}	4.66^{a}
365	−17.83	−1.87	−1.44	4.83	1.22^{a}	0.91^{c}	−3.07	−2.76	2.60	1.28	2.10^{c}	2.74	4.01^{b}	4.82^{c}
730	−5.10	−6.10	−0.87	6.36	1.00^{b}	0.46^{c}	−6.56	−16.86	2.02	1.32^{c}	1.70^{c}	2.28	4.43^{c}	6.89
All	−5.24	−0.41	−0.31	5.52^{a}	3.25^{a}	3.26^{a}	−0.03	−1.83	3.89^{a}	2.63^{a}	3.37^{a}	4.03^{a}	5.08^{a}	6.34^{a}
	Twenty days ahead (%)													
30	−3.05	−2.18	−5.37	−5.73	−6.14	−6.76	1.63	−2.43	−1.14	−1.99	−2.14	−0.92	2.06	2.60^{a}
91	−10.36	−4.93	−7.08	−4.53	−7.66	−8.71	−0.53	−3.65	−3.03	−4.19	−4.11	−2.81	0.52	1.04^{a}
152	−10.81	−3.38	−7.55	−5.47	−7.73	−9.08	0.38	−3.58	−2.95	−3.84	−3.88	−2.77	0.36	1.04^{c}
365	−15.93	−2.56	−6.81	−5.75	−7.22	−9.23	1.22	−3.33	−2.03	−3.21	−2.88	−1.91	1.78	1.29
730	−6.42	−2.45	−4.32	−5.76	−5.64	−7.96	−0.30	−6.37	−0.14	−2.07	−1.05	−0.10	4.48	1.91
All	−7.26	−3.06	−6.18	−5.41	−6.81	−7.86	0.78^{b}	−3.20	−1.89	−2.91	−2.88	−1.69	1.55^{b}	1.84^{a}

This table reports the R^2_{OS} statistics of the implied volatility forecasts of the 14 prediction models. AR(1) model is used as the benchmark to calculate the R^2_{OS} statistics. A positive R^2_{OS} statistic indicates that the prediction model outperforms the benchmark model. The statistical significance for the R^2_{OS} statistic is based on the p-value of the MSPE-adjusted statistic. The Hodrick (1992) standard error is used for the p-value calculation to account for the impact of overlapping residuals. [a, b, c]denote significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1996 to 2015, while the out-of-sample forecast starts from 2002 and ends in 2015.

5% significance level. Models VARC, ECM1, and ECM2 that perform well in the 1-day-ahead and 5-day-ahead forecasts fail to beat the benchmark model in the twenty-day-ahead forecast.

In order to visually observe the performance of the models over time, we also calculate their monthly aggregate out-of-sample forecast errors and compare them with those of the benchmark model. Figure 2 plots the difference of the monthly aggregate out-of-sample forecast errors between model VARC, one of the best-performing models reported in Table 6, and the benchmark model. A negative value means that model VARC performs better in that month. We standardize the series to make the pattern clear. Figure 2 shows that, for the 1-day-ahead forecast of all maturities implied volatilities and the 5-day-ahead forecast of short-maturity implied volatilities, most of the differences are negative, suggesting that model VARC consistently outperforms the benchmark model during the sample period.

3.2 | Economic significance

The statistical significance results in Table 4 suggest that these models can forecast implied volatilities rather well up to 5 days ahead. To explore the economic significance of this predictability, we further develop the option trading strategies as described in section 2. Following Goncalves and Guidolin (2006), we apply several filters to avoid microstructure-related bias. First, we exclude thinly traded options with less than 100 contracts per day. Second, we keep only the options with a positive bid-ask spread, a positive open interest and a positive implied volatility. Third, we exclude noisy contracts with fewer than 6 trading days to maturity and prices lower than \$3/8. Since the Leland's alphas are subtracted from those of the benchmark, any model with economically significant predictability returns a positive gain on Leland's alpha.

Table 5 reports the results. The performance of models VARC, ECM1, and ECM2 continues to be among the best. The combination forecasts also provide better economic performance than the benchmark model. The economic significance of the 1-day-ahead forecast is much stronger than that of the 5-day-ahead forecast. MidP uses mid price and does not assume any transaction cost. For model VARC, which performs the best, the gain on Leland's alpha for the 1-day-ahead forecast using MidP is 11.13% and significant at the 1% level. It declines to 2.13% for the 5-day-ahead forecast. In sharp contrast, none of the 14 models considered is economically significant for the 20-day-ahead forecast. This is consistent with our earlier finding that the historical implied volatility surface information is important for predicting the implied volatility only up to 1 week ahead.

The economic significance results are robust to the impact of transaction costs. Results change little when different levels of transaction costs are introduced. For example, the gain on Leland's alpha of 1-day-ahead forecast using VARC only decreases from 11.13% to 10.79% when we change the effective option spread from 0% (using MidP) to 100% of the quoted spread. One possible reason is that both the tested models and the benchmark model involve transaction costs. As a result the impact of transaction costs on their performance difference is balanced out.

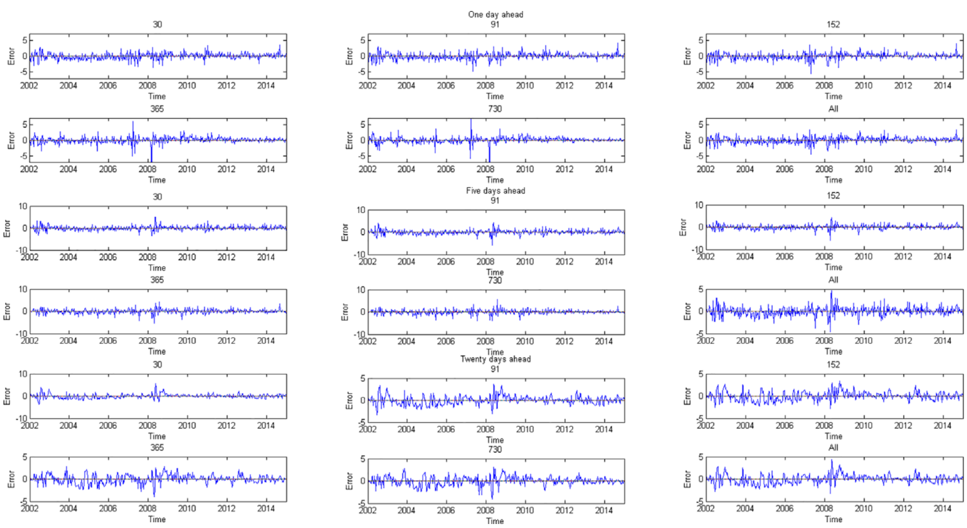


FIGURE 2 Difference of the out-of-sample forecast errors between the VAR(1) model on volatility change (VARC) and the benchmark model. This graph plots the standardized difference of monthly aggregate out-of-sample forecast errors between the best-performing model, VAR(1) on volatility change, and the benchmark model. A negative value means a smaller out-of-sample forecast error for the VARC model [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Economic significance of implied volatility forecast: Gain on Leland's alpha

	Model													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	NSAR	**NSVAR**	**VARL**	**VARC**	**ECM1**	**ECM2**	**PCA**	**EC**	**MC**	**MD**	**TM**	**DMSPE1**	**DMSPE2**	**MMA**
	One day ahead													
Mid price	−0.23	1.25	6.37^{a}	11.13a	9.86^{a}	10.57^{a}	−0.01	0.64	7.80^{a}	7.49^{a}	7.85^{a}	8.20^{a}	8.16^{a}	12.87^{a}
10% effective spread	−0.23	1.25	6.37^{a}	11.12^{a}	9.86^{a}	10.57^{a}	−0.01	0.64	7.80^{a}	7.49^{a}	7.85^{a}	8.20^{a}	8.16^{a}	12.87^{a}
25% effective spread	−0.23	1.25	6.38^{a}	11.11^{a}	9.86^{a}	10.57^{a}	−0.02	0.63	7.80^{a}	7.49^{a}	7.85^{a}	8.20^{a}	8.16^{a}	12.87^{a}
100% effective spread	−0.32	1.29	6.46^{a}	10.79^{a}	9.89	10.62^{a}	−0.21	0.51	7.81^{a}	7.54^{a}	7.88^{a}	8.21^{a}	8.19^{a}	12.82^{a}
	Five days ahead													
Mid price	−1.18	−1.00	−0.11	2.13^{c}	1.30	1.45	−0.05	−1.13	0.58	0.01	0.32	0.63	1.42^{c}	2.32^{c}
10% effective spread	−1.18	−1.00	−0.11	2.13^{c}	1.30	1.46	−0.05	−1.13	0.58	0.01	0.32	0.63	1.42^{c}	2.32^{c}
25% effective spread	−1.19	−0.99	−0.10	2.12^{c}	1.30	1.46	−0.06	−1.13	0.59	0.02	0.32	0.63	1.42^{c}	2.31^{c}
100% effective spread	−1.32	−0.89	0.06	1.85	1.38	1.55	−0.15	−1.16	0.67	0.14	0.42	0.71	1.51	2.23^{c}
	Twenty days ahead													
Mid price	−1.42	−1.33	−1.09	−1.39	−1.11	−1.47	−0.10	−1.34	−1.15	−1.23	−1.22	−1.14	−0.63	0.49
10% effective spread	−1.42	−1.33	−1.09	−1.40	−1.11	−1.47	−0.10	−1.34	−1.15	−1.23	−1.22	−1.15	−0.63	0.49
25% effective spread	−1.43	−1.32	−1.08	−1.41	−1.10	−1.46	−0.10	−1.34	−1.15	−1.22	−1.21	−1.14	−0.63	0.49
100% effective spread	−1.61	−1.19	−0.96	−1.74	−0.97	−1.29	−0.11	−1.24	−1.05	−1.11	−1.10	−1.06	−0.54	0.57

This table reports the gain on Leland's alpha of each model. On date t, we long (short) an option if the forecast volatility for that maturity at date $t + h$ is larger (smaller) than the current volatility. We consider options with maturities of 30, 91, 152, 365, and 730 days. Following Constantinides et al. (2013), we construct option portfolios targeting these maturities. We re-balance the portfolio daily and repeat the trade in the out-of-sample period. We delta-hedge our option portfolio. We first use the mid price (MidP) that does not assume any transaction costs to calculate the gain on Leland's alpha. We then assume the effective option spread to be 10%, 25%, and 100% of quoted spread. $^{a, b, c}$denote significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1996 to 2015, while the out-of-sample forecast starts from 2002 and ends in 2015.

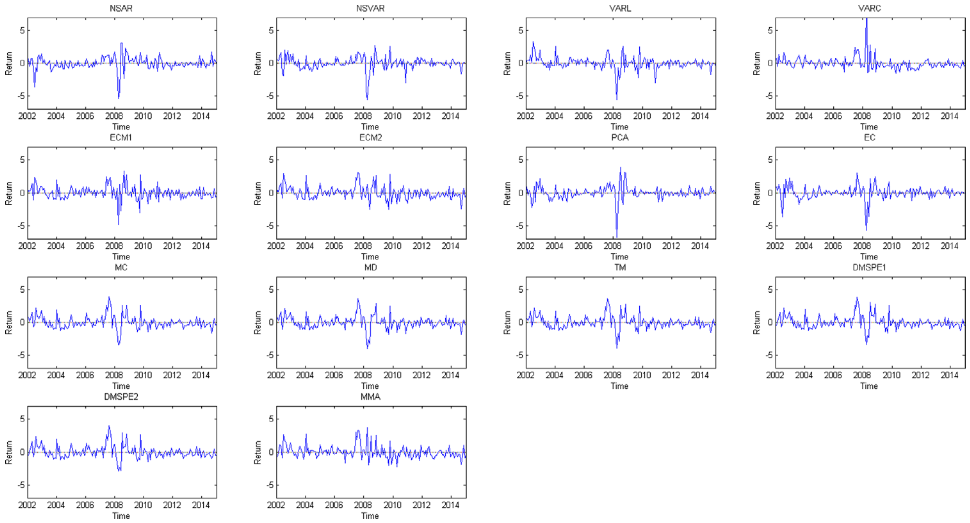


FIGURE 3 Time series of monthly portfolio return: 1-day-ahead forecast. This graph plots the monthly return of portfolios that are based on the 1-day-ahead forecast of implied volatility by 14 different models [Color figure can be viewed at wileyonlinelibrary.com]

Figure 3 plots the standardized aggregate monthly returns of the portfolios that are based on the forecast one day ahead. For those models that have significantly positive gain on Leland's alpha (models VARL, VARC, ECM1, ECM2, MC, TM, DMSPE1, DMSPE2, and MMA), returns are relatively stable during the normal time, but become volatile during the crisis period. Most have a large downward spike during the crisis, suggesting that these trading strategies could be subject to downside risk. The only exception is model VARC. It has a sudden return increase during the crisis, hence providing a better hedge against downside risk compared with other models. Figure 4 plots the standardized aggregate monthly returns of the portfolios that are based on the forecast five days ahead, and the findings are similar.

4 | ROBUSTNESS CHECKS

4.1 | Out-of-sample forecast during the recent financial crisis

Our data covers the recent financial crisis period. One question is whether the crisis has any impact on the implied volatility predictability. We examine the performance of out-of-sample forecast between December 2007 and June 2009, the recession period indicated by the National Bureau of Economic Research (NBER). Table 6 presents the results of the statistical and

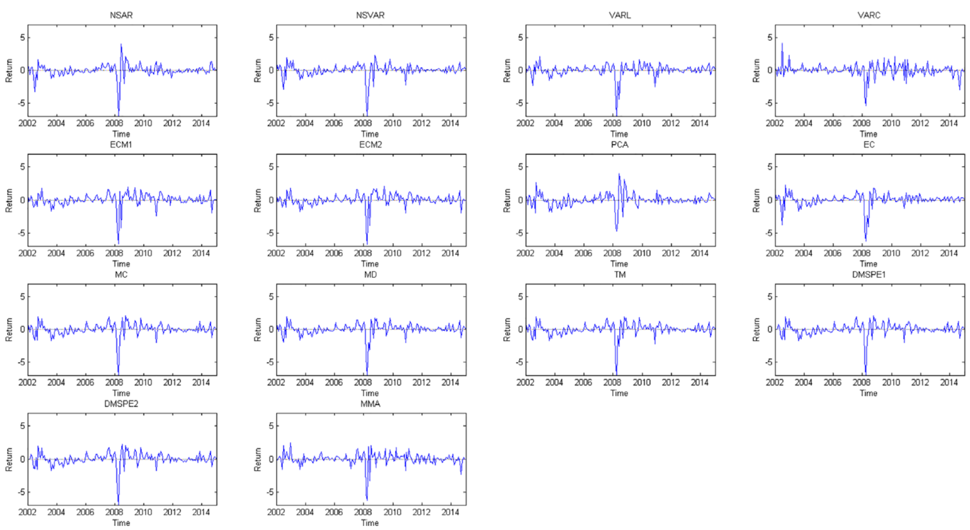


FIGURE 4 Time series of monthly portfolio return: 5-day-ahead forecast. This graph plots the monthly return of portfolios that are based on the five-day-ahead forecast of implied volatility by 14 different models [Color figure can be viewed at wileyonlinelibrary.com]

economic significance during this period. Panel A reports the R^2_{OS} statistics. In general, the results of 1-day-ahead forecasts during the financial crisis are close to those for the full sample period. Overall, there are nine models that outperform the benchmark model at the 5% significance level in predicting the implied volatility 1 day ahead. However, the results of 5-day-ahead and 20-day-ahead forecasts are weaker compared with the full sample period. Only two models are significant for the 5-day-ahead forecast, and none of the 20-day-ahead forecasts is significant. This result is different from the finding of stronger predictability during the recession period on stock market (Rapach et al., 2010) and corporate bond market (Lin et al., 2014). Rapach et al. (2010) and Lin et al. (2014) focus on risk premium forecast. They use macroeconomic variables and aggregate market variables to forecast the return at monthly or longer horizons, and impose the non-negative restriction to the forecast. Risk premium tends to be more predictable during crisis period. On the other hand, we are interested in how fast implied volatility reflects new information and focus on short horizon predictability using historical implied volatility surface information. We are testing a different question and the results are not comparable. One possible reason is that during a crisis period, investors are more sensitive to information available in the markets. As a result, it takes less time for the option market to absorb new information.

Panel B of Table 6 reports the economic significance results. Different from the statistical result, the economic significance of predictability actually strengthens. For example, the option trading strategy using 1-day-ahead forecasts based on model VARC generates a 33.90% gain on Leland's alpha. The gain on Leland's alpha slightly decreases to 31.18% when 100% effective option spread is used. They are much higher than those reported in Table 5. This shows that historical information is more economically important during a crisis period, which is consistent with Loh and Stulz (2014) who find that analysts tend to make poor forecasts during the crisis, but that the forecasts become more influential once the forecasts are adjusted.

4.2 | Out-of-sample forecast of put option

Another question is whether the findings using call options could be extended to put options. To answer this question, we run our tests using implied volatility surface of put options. We use the ATM put options and the put options with $\Delta = 0.40$ and $\Delta = 0.60$.

The top panel of Table 7 reports the R^2_{OS} statistics and the following panels report the gain on Leland's alpha of different horizons. For simplicity, we only present the results of all maturity and moneyness. The implied volatilities of put options are statistically predictable up to 20 days. Such predictability is stronger than that of call options. Model VARC, ECM1, ECM2, and combination forecasts continue to perform well.

The predictability of implied volatilities of put options is also of economic significance. For the one-day-ahead forecast, model VARC generates a gain on Leland's alpha of 12.05% if MidP is used, and 10.79% if 100% effective option spread is used. Both of them are significant at the 1% level. The economic significance of 5-day-ahead forecast is still strong. Model VARC generates a gain on Leland's alpha of 5.16% if MidP is used, and 1.85% if 100% effective option spread is used. They are stronger than those of call options. Overall, the results using put option data strengthen our findings that implied volatility surface contains useful information for the forecast of implied volatilities.

4.3 | Out-of-sample forecast with other benchmark

Recently, Egloff et al. (2010) and Johnson (2017) show that slope is an important predictor of implied variance. As a robustness check, we replace the benchmark model of AR(1) with the two-factor model that uses level and slope, and re-run all the tests. Unreported results show that the earlier well-performing models, such as VARC, ECM1, ECM2, and combination forecast, have significantly positive R^2_{OS} up to 1 week.[13] The overall R^2_{OS} of VARC model is 12.47% for the 1-day ahead forecast, and is significant at the 1% level. The results of ECM1, ECM2 and combination forecast are similar and also significant at the 1% level. These models lose predictive power after 1 week.

Results of economic significance analysis also suggest that VARC, ECM1, ECM2, and combination forecast outperform the two-factor benchmark model. The gain on Leland's alpha of the 1-day ahead forecast using VARC is 6.11% and significant at the 1% level. Results continue to be significant for the five-day-ahead forecast, and become insignificant after 1 week.

Our empirical analysis suggests that for the short-horizon implied volatility forecast, a flexible model framework such as VARC and others is able to use the surface information more efficiently, and provides a better out-of-sample result. On the other hand, when the forecast horizon is beyond 1 week, more flexible models generate more noise, and simple models such as AR and two-factor model start to function better. This finding has useful implication for the portfolio management.

[13]The results are available upon request.

TABLE 6 Predictability of implied volatility during the financial crisis

	Model													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Maturity days	NSAR	NSVAR	VARL	VARC	ECM1	ECM2	PCA	EC	MC	MD	TM	DMSPE1	DMSPE2	MMA
Panel A. R^2_{OS}														
	One day ahead (%)													
30	−2.71	−0.92	−1.50	7.66[a]	3.71[a]	4.22[a]	0.07	−0.33	4.23[a]	2.96[a]	3.59[a]	4.36[a]	4.30[a]	6.28[a]
91	−7.48	−1.15	−0.05	7.87[a]	3.73[a]	5.15	−1.81	−3.07	4.08[a]	3.37[a]	3.65[a]	4.36[a]	3.96[a]	7.31[a]
152	−10.87	−4.46	−0.46	7.83[a]	3.74[a]	5.25[a]	−5.76	−7.30	2.73[a]	1.46[b]	2.25[a]	3.11[a]	2.82[a]	7.54[a]
365	−36.33	−9.01	−0.78	9.36[a]	2.93[a]	4.45[a]	−7.04	−4.62	4.11[a]	1.98[b]	3.40[a]	4.70[a]	4.42[a]	9.75[a]
730	−7.46	−11.08	−3.14	11.37[a]	−3.36	−2.44	−14.30	−53.58	1.84[a]	−0.57	1.00[a]	2.77[a]	3.97[a]	12.94[a]
All	−6.90	−2.21	−1.05	7.94[a]	3.45[a]	4.36[a]	−1.97	−3.81	3.92[a]	2.69[a]	3.34[a]	4.17[a]	4.03[a]	7.09[a]
	Five days ahead (%)													
30	−0.54	−1.26	−10.48	6.08[a]	−3.37	−1.63	0.98	−3.10	0.97	−1.46	−0.21	1.06	1.92	4.56
91	−8.12	−6.19	−10.81	3.90[b]	−3.79	−2.11	−0.98	−6.53	−1.17	−4.12	−2.05	−1.11	−0.12	2.99
152	−7.22	−4.28	−11.02	4.39[b]	−4.24	−2.59	0.51	−6.11	−0.35	−3.10	−1.37	−0.34	0.90	3.54
365	−23.44	−7.57	−12.64	3.64[c]	−6.32	−6.62	−0.51	−7.49	−1.34	−5.86	−2.97	−1.39	0.37	1.71
730	−6.20	−9.37	−10.76	2.04	−5.85	−8.26	−5.16	−24.68	−0.87	−5.31	−2.53	−0.85	2.79[c]	2.55
All	−4.67	−3.41	−10.77	5.09[a]	−3.83	−2.37	0.18	−5.24	0.11	−2.66	−1.02	0.18	1.26[c]	3.84[a]
	Twenty days ahead (%)													
30	−2.06	−6.05	−19.86	−15.71	−17.21	−18.72	−1.71	−7.70	−8.99	−9.42	−9.94	−8.76	−6.90	−5.58
91	−6.33	−7.46	−18.34	−9.32	−17.30	−18.98	−0.08	−7.53	−8.48	−10.44	−10.14	−8.31	−6.50	−4.04
152	−7.14	−6.32	−18.58	−8.47	−18.35	−20.47	1.92	−7.59	−8.26	−10.50	−10.15	−8.07	−6.10	−2.64
365	−15.42	−6.35	−17.10	−4.07	−19.80	−22.41	5.50	−7.79	−7.30	−9.92	−9.38	−7.16	−4.45	0.58
730	−5.74	−0.59	−7.98	−2.38	−13.03	−14.76	10.93[a]	−2.31	0.16	−3.99	−2.40	0.21	4.37	7.57[c]
All	−4.98	−6.21	−18.58	−11.67	−17.41	−19.14	0.31	−7.41	−8.24	−9.62	−9.65	−8.04	−6.01	−3.75

(Continues)

TABLE 6 (Continued)

	Model													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Maturity days	**NSAR**	**NSVAR**	**VARL**	**VARC**	**ECM1**	**ECM2**	**PCA**	**EC**	**MC**	**MD**	**TM**	**DMSPE1**	**DMSPE2**	**MMA**
Panel B. Economic significance: Gain on Leland's alpha														
	One day ahead													
MidP	7.33[b]	2.10	4.48	33.90[b]	19.52	24.87[c]	7.21	4.18	19.71[b]	15.57	19.23[c]	21.44[b]	19.88[b]	28.84[b]
10% effective spread	7.32[b]	2.09[b]	4.45[b]	33.87[b]	19.49	24.84[b]	7.19[b]	4.16[b]	19.69[b]	15.55[b]	19.21[b]	21.42[b]	19.86[b]	28.81
25% effective spread	7.29[b]	2.01[b]	4.31[b]	33.73[b]	19.35[b]	24.70[b]	7.07[b]	4.04[b]	19.59[b]	15.44[b]	19.11[b]	21.32[b]	19.76[b]	28.68[b]
100% effective spread	6.77[b]	0.57[b]	1.74[b]	31.18[b]	16.78[b]	22.24[b]	5.02[b]	1.94[b]	17.90[b]	13.56[b]	17.32[b]	19.51[b]	18.00[b]	26.25[b]
	Five days ahead													
MidP	1.19	−4.36	−6.11	−1.92	1.23	2.76	4.57	−4.52	1.82	−3.14	−0.50	1.33	4.03	0.26
10% effective spread	1.18	−4.37	−6.12	−1.95	1.21	2.75	4.56	−4.53	1.81	−3.15	−0.52	1.31	4.01	0.24
25% effective spread	1.17	−4.42	−6.21	−2.07	1.11	2.65	4.48	−4.61	1.75	−3.22	−0.59	1.25	3.95	0.13
100% effective spread	0.87	−5.33	−7.73	−4.23	−0.66	0.98	3.03	−6.05	0.59	−4.39	−1.84	0.06	2.78	−1.83
	Twenty days ahead													
MidP	1.32	−0.29	−4.28	−6.49	−2.34	−3.74	1.83	−2.71	−0.05	−1.03	−0.95	0.00	0.95	1.92
10% effective spread	1.32	−0.30	−4.29	−6.51	−2.35	−3.75	1.82	−2.72	−0.05	−1.04	−0.95	−0.01	0.94	1.91
25% effective spread	1.29	−0.45	−4.51	−7.01	−2.55	−3.93	1.61	−2.92	−0.20	−1.20	−1.12	−0.16	0.79	1.67
100% effective spread	1.17	−0.91	−5.20	−8.56	−3.19	−4.51	0.94	−3.56	−0.67	−1.72	−1.63	−0.63	0.32	0.89

This table reports the predictability of implied volatility during the financial crisis period. Panel A reports the R_{OS}^2 statistics, while Panel B reports the economic significance results. AR(1) model is used as the benchmark to calculate the R_{OS}^2 statistics. A positive R_{OS}^2 statistic indicates that the prediction model outperforms the benchmark model. The statistical significance for the R_{OS}^2 statistic is based on the p-value of the MSPE-adjusted statistic. The Hodrick (1992) standard error is used for the p-value calculation to account for the impact of overlapping residuals. Panel B reports the gain on Leland's alpha of each model during the recent financial crisis. On date t, we long (short) an option if the forecast volatility for that maturity at date $t+h$ is larger (smaller) than the current volatility. We consider options with maturities of 30, 91, 152, 365, and 730 days. Following Constantinides et al. (2013), we construct option portfolios targeting these maturities. We rebalance the portfolio daily and repeat the trade in the out-of-sample period. We delta-hedge our option portfolio. We first use the mid price (MidP) that does not assume any transaction costs to calculate the gain on Leland's alpha. We then assume the effective option spread to be 10%, 25%, and 100% of quoted spread. [a, b, c] denote significance at the 1%, 5% and 10% level, respectively. The financial crisis period is between December 2007 and June 2009.

TABLE 7 Out-of-sample forecast of implied volatilities of put options

	Model	(1) NSAR	(2) NSVAR	(3) VARL	(4) VARC	(5) ECM1	(6) ECM2	(7) PCA	(8) EC	(9) MC	(10) MD	(11) TM	(12) DMSPE1	(13) DMSPE2	(14) MMA
	One day ahead	−7.70	−1.97	1.12[a]	3.32[a]	2.71[a]	2.72[a]	−6.51	−7.45	2.29[a]	2.17[a]	2.32[a]	2.54[a]	2.67[a]	3.77[a]
R^2_{OS}	Five days ahead	−6.26	−0.58	−0.12	4.66[a]	2.83[a]	2.74[a]	−0.63	−2.26	2.92[a]	2.04[a]	2.39[a]	3.07[a]	4.10[a]	5.38[a]
	Twenty days ahead	−5.38	−3.46	−5.57	−3.71	−5.79	−6.44	0.79[a]	−3.77	−1.12	−2.36	−2.20	−1.03	2.42[a]	2.56[a]
		One day ahead													
	MidP	1.00	2.41	9.83[a]	12.05[a]	10.27[a]	9.65[a]	3.20[b]	2.78[c]	6.13[a]	7.48[a]	6.39[a]	6.31[a]	6.65[a]	11.97[a]
Gain on	10% effective spread	1.00	2.41	9.83	12.05[a]	10.27[a]	9.65[a]	3.20[b]	2.78[c]	6.13[a]	7.48[a]	6.39[a]	6.31[a]	6.65[a]	11.97[a]
Leland's alpha	25% effective spread	1.01	2.40	9.85[a]	12.07[a]	10.27[a]	9.64[a]	3.20[b]	2.77[c]	6.13[a]	7.48[a]	6.39[a]	6.31[a]	6.65[a]	11.98[a]
	100% effective spread	−0.32	1.29	6.46[a]	10.79[a]	9.89[a]	10.62[a]	−0.21	0.51	7.81[a]	7.54[a]	7.88[a]	8.21[a]	8.19[a]	12.82[a]
		Five days ahead													
	MidP	0.23	0.57	3.00[a]	5.16[a]	3.84[a]	3.76[a]	1.89[b]	1.49	2.15[b]	1.85[b]	2.19	2.23[b]	3.11[a]	5.23[a]
Gain on	10% effective spread	0.23	0.57	3.00[a]	5.16[a]	3.84[a]	3.76[a]	1.89[b]	1.48	2.15[b]	1.85[b]	2.19[b]	2.23[b]	3.10[a]	5.23[a]
Leland's alpha	25% effective spread	1.01	2.40	9.85[a]	12.07[a]	10.27[a]	9.64[a]	3.20[b]	2.77	6.13[a]	7.48[a]	6.39[a]	6.31[a]	6.65[a]	11.98[a]
	100% effective spread	−1.32	−0.89	0.06	1.85	1.38	1.55	−0.15	−1.16	0.67	0.14	0.42	0.71	1.51[c]	2.23[c]
		Twenty days ahead													
	MidP	−1.00	0.08	0.90	0.51	0.61	−0.20	0.67	0.28	−0.21	0.16	0.04	−0.19	0.70	0.73
Gain on	10% effective spread	−0.99	0.07	0.90	0.51	0.61	−0.20	0.67	0.28	−0.21	0.16	0.04	−0.19	0.69	0.72
Leland's alpha	25% effective spread	−0.97	0.06	0.90	0.52	0.60	−0.21	0.66	0.26	−0.22	0.15	0.03	−0.21	0.68	0.71
	100% effective spread	−1.61	−1.19	−0.96	−1.74	−0.97	−1.29	−0.11	−1.24	−1.05	−1.11	−1.10	−1.06	−0.54	0.57

This table reports the out-of-sample forecast of implied volatilities of put options by the 14 models. The top panel report the results of R^2_{OS}, while the following panels reports the results of the gain on Leland's alpha at different horizons. For simplicity, we only report the results of all maturity and moneyness. [a, b, c] denote significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1996 to 2015, while the out-of-sample forecast starts from 2002 and ends in 2015.

4.4 | Option data with a different Δ range

We use options with Δ between 0.40 and 0.60 in our empirical analysis. Literature shows that the most liquid options are ATM as well as +0.25 and −0.25 Δ options, which also contain valuable information.[14] To test whether our results are robust to the use of data with a different Δ, we re-run our analysis using the option data with Δ between 0.30 and 0.70, which covers the moneyness between 25% and 75%.

Table 8 reports the results. Panel A and Panel B report the result of call and put options, respectively. We report the overall R^2_{OS} and the gain on Leland's alpha for the forecast of 1 day ahead, 5 days ahead, and 20 days ahead. Results clearly show that implied volatility is still predictable when the data with Δ between 0.30 and 0.70 are used. The results are close to those using the data with Δ between 0.40 and 0.60. For the 1-day-ahead forecast of call options, there are ten models with significant R^2_{OS} and gain on Leland's alpha. For the 5-day-ahead forecast of call option, there are nine models with significant R^2_{OS} and three models with significant gain on Leland's alpha. Results become insignificant for the 20-day-ahead forecast. The results of put option in Panel B show a similar pattern.

4.5 | Gain on alpha from a different asset pricing model

We use the gain on Leland's alpha as the economic significance measure. Leland's alpha only considers the impact of market risk on the option portfolio return. In order to test whether the economic significance is robust to the choice of asset pricing model, we run the regression of Chen, Roll, and Ross (1986) five factor model on long-short option portfolio return to get the gain on alpha. We construct long-short option portfolios following the option trading strategies as described in section 2 for each model. We calculate the monthly cumulative option portfolio return of each predictive model, and then the difference of the option portfolio return between the predictive model and the benchmark model. We then run the time series regression of the option portfolio return difference on the five factors of Chen et al. (1986),

$$r_{i,t} - r_{0,t} = \mathrm{Alpha} + \beta_{\mathrm{MP}}\mathrm{MP}_t + \beta_{\mathrm{DEI}}\mathrm{DEI}_t + \beta_{\mathrm{UI}}\mathrm{UI}_t + \beta_{\mathrm{UPR}}\mathrm{UPR}_t + \beta_{\mathrm{UTS}}\mathrm{UTS}_t + \varepsilon_{\mathrm{it}} \tag{10}$$

where r_{it} is the option portfolio return of predictive model i in month t, r_{0t} is the option portfolio return of the benchmark model in month t, MP_t, DEI_t, UI_t, UPR_t, and UTS_t are industrial production growth, changes in expected inflation, unexpected inflation, risk premium, and term structure factor in month t, respectively.[15] We are interested in whether the intercept, Alpha, is significant after controlling for the five factors.

Table 9 reports the results of 1-day-ahead forecast. We report the results of call and put options with Δ between 0.40 and 0.60 (upper panel) and between 0.30 and 0.70 (bottom panel). Results strongly show that implied volatility predictability is economically significant after controlling for the five factors of Chen et al. (1986). For the call option with Δ between 0.40 and 0.60, there are 10 models with significant gains on alpha if mid-price is used to calculate the option return. Results change little if 100% effective spread is used. Results of put option are stronger than those of call option. Using data with Δ between 0.30 and 0.70 generates a similar pattern. These results suggest that there exist significant economic gains of implied volatility predictability form using the information of historical implied volatility surface.

5 | STOCHASTIC VOLATILITY MODEL

Our finding that implied volatility surface contains information for the prediction of implied volatilities is consistent with the literature of multi-factor stochastic volatility model. To test this hypothesis, we calibrate the two-factor stochastic volatility option pricing model (9) to the implied volatility data. Following Christoffersen et al. (2009) we calibrate the option pricing formula to the weekly data of ATM calls and puts.[16]

Figure 5 plots the time series of the two variance factors for the calls (the upper panel) and puts (the bottom panel). The variance factors fluctuate a lot over time, and reach a peak during the crisis period. The first factor is much more persistent than

[14]See for example, Carr and Wu (2007).

[15]MP_t, DEI_t and UI_t are obtained from Federal Reserve Bank of St Louis, while UPR_t and UTS_t are downloaded from Amit Goyal's website.

[16]Appendix 2 reports the calibration results in each year for calls and puts respectively.

TABLE 8 Predictability of implied volatility with Δ between 0.30 and 0.70

		Model													
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
		NSAR	NSVAR	VARL	VARC	ECM1	ECM2	PCA	EC	MC	MD	TM	DMSPE1	DMSPE2	MMA
Panel A. Call options															
	One day ahead	−8.00	−0.78	7.92[a]	10.32[a]	8.88[a]	9.22[a]	−5.37	−11.20	9.75[a]	10.09[a]	9.98[a]	10.50[a]	10.72[a]	13.34[a]
R^2_{OS}	Five days ahead	−5.59	−0.46	2.54	7.20[a]	5.16[a]	5.19[a]	−0.40	−4.04	6.18[a]	5.73[a]	5.89[a]	6.43[a]	7.77[a]	8.68[a]
	Twenty days ahead	−7.74	−3.29	−5.73	−6.35	−6.07	−6.40	0.74	−3.39	−0.77	−1.94	−1.58	−0.55	3.19[c]	2.45[b]
		One day ahead													
	Mid price	0.08	1.20	10.79[a]	10.92[a]	12.80[a]	12.92[a]	−0.15	1.46	8.25[a]	9.23[a]	8.99[a]	9.33[a]	9.50[a]	14.27[a]
Gain on	10% effective spread	0.08	1.20	10.79[a]	10.92[a]	12.80[a]	12.92[a]	−0.15	1.46	8.25[a]	9.23[a]	8.99[a]	9.33[a]	9.50[a]	14.27[a]
Leland's alpha	25% effective spread	0.05	1.20	10.77[a]	10.84[a]	12.77[a]	12.91[a]	−0.21	1.40	8.22[a]	9.21[a]	8.96[a]	9.30[a]	9.47[a]	14.26[a]
	100% effective spread	−0.03	1.19	10.72[a]	10.57[a]	12.70[a]	12.85[a]	−0.39	1.22	8.13[a]	9.15[a]	8.89[a]	9.22[a]	9.40[a]	14.25[a]
		Five days ahead													
	Mid price	−0.87	−1.02	0.54	2.71[b]	1.41	1.52	0.16	−0.74	0.57	0.56	0.48	0.59	1.66[b]	2.94[b]
Gain on	10% effective spread	−0.87	−1.02	0.54	2.71[b]	1.41	1.52	0.16	−0.74	0.57	0.56	0.48	0.59	1.66[b]	2.94[b]
Leland's alpha	25% effective spread	−0.88	−1.02	0.54	2.70[b]	1.41	1.52	0.15	−0.75	0.57	0.56	0.48	0.59	1.66[b]	2.93[b]
	100% effective spread	−1.02	−0.96	0.57	2.43[c]	1.42	1.52	0.00	−0.90	0.54	0.57	0.47	0.56	1.64[b]	2.88[b]
		Twenty days ahead													
	Mid price	−1.17	−1.30	−0.71	−1.26	−1.03	−1.12	−0.21	−1.55	−1.05	−1.11	−1.12	−1.07	−0.20	0.17
Gain on	10% effective spread	−1.17	−1.30	−0.71	−1.27	−1.03	−1.12	−0.21	−1.55	−1.05	−1.11	−1.12	−1.07	−0.20	0.17
Leland's alpha	25% effective spread	−1.18	−1.29	−0.70	−1.29	−1.02	−1.12	−0.21	−1.55	−1.05	−1.10	−1.12	−1.07	−0.20	0.17
	100% effective spread	−1.37	−1.18	−0.65	−1.61	−0.93	−1.01	−0.30	−1.60	−0.98	−1.05	−1.04	−1.01	−0.16	0.22

(Continues)

TABLE 8 (Continued)

		Model													
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
		NSAR	NSVAR	VARL	VARC	ECM1	ECM2	PCA	EC	MC	MD	TM	DMSPE1	DMSPE2	MMA
Panel B. Put options															
	One day ahead	−13.75	−2.85	4.57[a]	5.47[a]	4.96[a]	5.60[a]	−8.41	−11.83	4.91[a]	5.72[a]	5.50[a]	5.43[a]	5.30[a]	6.19[a]
R^2_{OS}	Five days ahead	−12.58	−3.78	1.84[a]	5.74[a]	2.75[a]	2.68[a]	−1.38	−6.13	2.95[a]	2.35[a]	2.64[a]	3.30[a]	5.00[a]	6.82[a]
	Twenty days ahead	−7.82	−8.22	−14.00	−2.81	−14.97	−16.06	0.04	−8.82	−5.49	−8.02	−7.30	−5.27	−1.23	0.69[b]
		One day ahead													
	Mid price	0.22	1.82	9.63[a]	11.41[a]	9.10[a]	8.89[a]	2.91[a]	2.82[c]	5.95[a]	7.77[a]	6.18[a]	5.79[a]	6.40[a]	11.39[a]
Gain on	10% effective spread	0.23	1.82	9.63[a]	11.41[a]	9.10[a]	8.89[a]	2.91[c]	2.82[c]	5.95[a]	7.77[a]	6.18[a]	5.79[a]	6.40[a]	11.39[a]
Leland's alpha	25% effective spread	0.24	1.82	9.66[a]	11.43[a]	9.11[a]	8.89[a]	2.91[c]	2.82[c]	5.96[a]	7.78[a]	6.19[a]	5.80[a]	6.41[a]	11.39[a]
	100% effective spread	0.49	1.77	10.19[a]	11.75[a]	9.34[a]	9.02[a]	2.96[b]	2.84[c]	6.02[a]	7.95[a]	6.33[a]	5.92[a]	6.51[a]	11.45[a]
		Five days ahead													
	Mid price	0.05	0.14	2.16[c]	4.16[a]	2.29[c]	2.21[c]	2.50[a]	2.09[b]	1.50[c]	1.92[b]	1.64[c]	1.63[c]	2.97[a]	4.81[a]
Gain on	10% effective spread	0.05	0.14	2.16[c]	4.16[a]	2.29[c]	2.21[c]	2.50[a]	2.09[b]	1.50[c]	1.92[b]	1.64[c]	1.63[c]	2.97[a]	4.81[a]
Leland's alpha	25% effective spread	0.07	0.12	2.18[c]	4.18[a]	2.30[c]	2.22[c]	2.50[a]	2.08[b]	1.51[c]	1.92[b]	1.64[c]	1.63[c]	2.97[a]	4.81[a]
	100% effective spread	0.37	−0.11	2.47[c]	4.51[a]	2.37[c]	2.27[c]	2.50[a]	2.01[b]	1.53[c]	1.94[b]	1.68[c]	1.67[c]	3.00[a]	4.87[a]
		Twenty days ahead													
	Mid price	−0.62	−0.04	0.49	1.27	0.18	−0.02	0.62	0.21	0.18	0.02	0.14	0.16	1.68[a]	1.97[a]
Gain on	10% effective spread	−0.62	−0.04	0.50	1.28	0.18	−0.02	0.62	0.21	0.18	0.02	0.14	0.16	1.68[a]	1.97[a]
Leland's alpha	25% effective spread	−0.60	−0.05	0.52	1.29	0.19	−0.01	0.61	0.20	0.18	0.03	0.14	0.15	1.67[a]	1.97[a]
	100% effective spread	−0.23	−0.30	0.84	1.62	0.35	0.09	0.54	0.10	0.18	0.04	0.12	0.15	1.63[a]	2.04[a]

This table reports the predictability of the implied volatility with Δ between 0.30 and 0.70. Panel A and B report the result of call and put options respectively. AR(1) model is used as the benchmark to calculate the R^2_{OS} statistics and gain on Leland's alpha. A positive R^2_{OS} statistic indicates that the prediction model outperforms the benchmark model. The statistical significance for the R^2_{OS} statistic is based on the p-value of the MSPE-adjusted statistic. The Hodrick (1992) standard error is used for the p-value calculation to account for the impact of overlapping residuals. Gain on Leland's alpha is used to measure the economic significance. On date t, we long (short) an option if the forecast volatility for that maturity at date $t+h$ is larger (smaller) than the current volatility. We consider options with maturities of 30, 91, 152, 365, and 730 days. Following Constantinides et al. (2013), we construct option portfolios targeting these maturities. We rebalance the portfolio daily and repeat the trade in the out-of-sample period. We delta-hedge our option portfolio. We first use the mid price (MidP) that does not assume any transaction costs to calculate the gain on Leland's alpha. We then assume the effective option spread to be 10%, 25%, and 100% of quoted spread. [a, b, c]denote significance at the 1%, 5%, and 10% level, respectively.

TABLE 9 Gain on alpha of option portfolio return using Chen et al. (1986) five factor model

		Model													
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
		NSAR	**NSVAR**	**VARL**	**VARC**	**ECM1**	**ECM2**	**PCA**	**EC**	**MC**	**MD**	**TM**	**DMSPE1**	**DMSPE2**	**MMA**
		Options with Δ between 0.40 and 0.60													
	Mid price	0.29	1.01	6.31[b]	13.28[a]	10.78[a]	12.24[a]	1.53	2.03	7.33[a]	7.08[a]	7.53[a]	7.84[a]	7.75[a]	14.74[a]
	10% effective spread	0.29	1.01	6.31[b]	13.28[a]	10.78[a]	12.23[a]	1.52	2.02	7.33[a]	7.07[a]	7.53[a]	8.04[a]	7.94[a]	14.93[a]
Call	25% effective spread	0.29	0.99	6.29[b]	13.25[a]	10.76[a]	12.22[a]	1.49	2.00	7.31[a]	7.06[a]	7.51[a]	8.31[a]	8.21[a]	15.20[a]
	100% effective spread	0.17	0.80	6.01[b]	12.67[a]	10.41[a]	11.94[a]	1.02	1.55	7.07[a]	6.83[a]	7.27[a]	9.51[a]	9.43[a]	16.29[a]
	Mid price	2.13	4.27[b]	14.33[a]	14.74[a]	13.13[a]	12.17[a]	5.79[a]	5.08[b]	8.61[a]	10.22[a]	8.65[a]	8.87[a]	9.19[a]	15.13[a]
	10% effective spread	2.14	4.27[b]	14.34[a]	14.74[a]	13.14[a]	12.17[a]	5.79[a]	5.09[b]	8.62[a]	10.22	8.65[a]	9.08[a]	9.40[a]	15.35[a]
Put	25% effective spread	2.14	4.27[b]	14.38[a]	14.78[a]	13.16[a]	12.19[a]	5.80[a]	5.10[b]	8.62[a]	10.23[a]	8.67[a]	9.42[a]	9.73[a]	15.70[a]
	100% effective spread	2.24	4.23[b]	15.14[a]	15.41[a]	13.60[a]	12.49[a]	5.99[a]	5.33[a]	8.80[a]	10.46[a]	8.89[a]	11.21[a]	11.50[a]	17.76[a]
		Options with Δ between 0.30 and 0.70													
	Mid price	0.64	0.83	12.15[a]	13.58[a]	14.86[a]	14.77[a]	0.87	3.67[c]	9.21[a]	9.12[a]	9.51[a]	10.05[a]	9.94[a]	16.00[a]
	10% effective spread	0.64	0.82	12.14[a]	13.57[a]	14.86[a]	14.77[a]	0.87	3.66[c]	9.21[a]	9.12[a]	9.51[a]	10.23[a]	10.13[a]	16.19
Call	25% effective spread	0.63	0.81	12.12[a]	13.54[a]	14.83[a]	14.74[a]	0.84	3.63[c]	9.19[a]	9.09	9.49[a]	10.50[a]	10.39[a]	16.46
	100% effective spread	0.47	0.58	11.66[a]	12.91[a]	14.32[a]	14.30[a]	0.33	3.03	8.80[a]	8.70a	9.09[a]	11.57[a]	11.44[a]	17.55[a]
	Mid price	1.24	3.85[b]	14.25[a]	15.32[a]	13.28[a]	12.55[a]	4.99[a]	5.74[a]	8.73[a]	11.19[a]	9.09[a]	8.71[a]	9.36[a]	14.57[a]
	10% effective spread	1.24	3.85[b]	14.26[a]	15.33[a]	13.28	12.55[a]	4.99[a]	5.75[a]	8.73[a]	11.19[a]	9.09[a]	8.93[a]	9.58[a]	14.78[a]
Put	25% effective spread	1.25	3.85[b]	14.32[a]	15.37[a]	13.32[a]	12.58[a]	5.00[a]	5.76[a]	8.74[a]	11.21[a]	9.11[a]	9.27[a]	9.91[a]	15.12[a]
	100% effective spread	1.48	3.91[b]	15.30[a]	16.00[a]	14.00[a]	13.08[a]	5.25[a]	6.09[a]	8.99[a]	11.64[a]	9.46[a]	11.19[a]	11.82[a]	17.10[a]

This table reports the gain on alpha of long-short option portfolio return based on the five factor model of Chen et al. (1986). We construct long-short option portfolios following the option trading strategies as described in section 2 for each model. We calculate the monthly cumulative option portfolio return of each predictive model, and then the difference of the option portfolio return between the predictive model and the benchmark model. We then run the time series regression of the option portfolio return difference on the five factors of Chen et al. (1986).

$r_{i,t} - r_{0,t} = \text{Alpha} + \beta_{\text{MP}}\text{MP}_t + \beta_{\text{DEI}}\text{DEI}_t + \beta_{\text{UI}}\text{UI}_t + \beta_{\text{UPR}}\text{UPR}_t + \beta_{\text{UTS}}\text{UTS}_t + \varepsilon_{\text{it}}$ where r_{it} is the option portfolio return of predictive model i in month t, r_{0t} is the option portfolio return of the benchmark model in month t, MP_t, DEI_t, UI_t, UPR_t, and UTS_t are industrial production growth, changes in expected inflation, unexpected inflation, risk premium and term structure factor in month t respectively. We consider options with maturities of 30, 91, 152, 365, and 730 days. Following Constantinides et al. (2013), we construct option portfolios targeting these maturities. We rebalance the portfolio daily and repeat the trade in the out-of-sample period. We delta-hedge our option portfolio. We first use the mid price (MidP) that does not assume any transaction costs to calculate the option return. We then assume the effective option spread to be 10%, 25%, and 100% of quoted spread. [a, b, c]denote significance at the 1%, 5%, and 10% level, respectively.

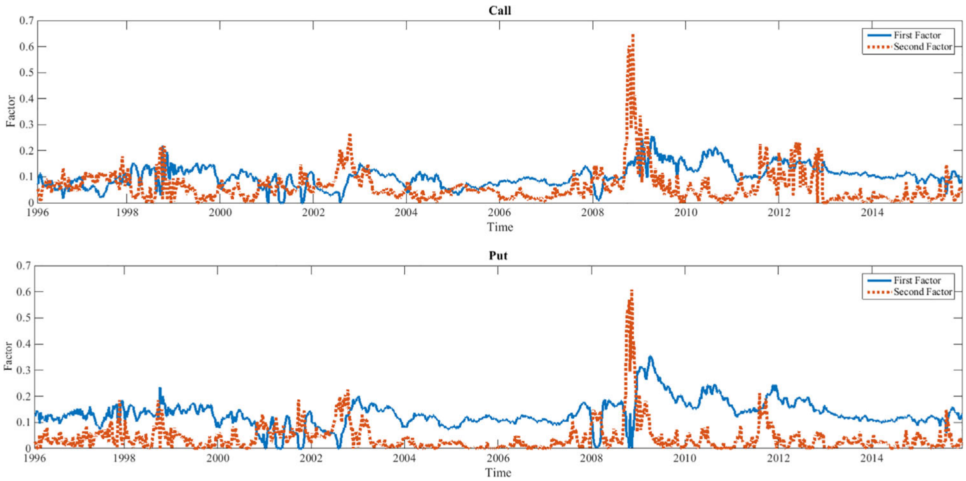


FIGURE 5 Extracted variance factors. This graph plots the time series of two extracted variance factors calibrated from the two-factor stochastic volatility option pricing model. The top panel plots the time series of two extracted variance factors for ATM calls, while the bottom panel plots the time series for ATM puts [Color figure can be viewed at wileyonlinelibrary.com]

the second factor. Indeed, the mean β values of calls (puts) are 0.21 (0.13) and 3.21 (1.70) for the first factor and the second factor, respectively. We therefore call the first factor the long-term variance factor while the second factor is the short-term variance factor.

We then run the univariate regressions

$$\sigma_t^2(\tau) = \alpha(\tau) + \beta_1(\tau)V_{1t} + \varepsilon_t(\tau), \tag{11}$$

TABLE 10 Relationship between implied volatility and the extracted long-term and short-term variance factors

	ATM call					ATM put				
	Long-term factor		Short-term factor			Long-term factor		Short-term factor		
Maturity (days)	β_1	*t*-stats	β_2	*t*-stats	Adj. R^2	β_1	*t*-stats	β_2	*t*-stats	Adj. R^2
30	0.25	4.09			0.06	0.09	1.26			0.01
			0.54	13.31	0.72			0.67	28.12	0.83
	0.23	7.92	0.53	14.83	0.78	0.24	14.40	0.71	32.55	0.92
91	0.28	5.97			0.13	0.16	2.82			0.06
			0.39	14.45	0.62			0.49	17.11	0.70
	0.27	10.36	0.39	15.93	0.74	0.27	15.61	0.54	22.05	0.88
182	0.30	8.34			0.23	0.19	4.25			0.12
			0.29	12.60	0.51			0.36	12.17	0.57
	0.29	13.33	0.29	13.19	0.72	0.28	17.22	0.41	16.05	0.83
365	0.31	10.63			0.33	0.19	5.16			0.17
			0.22	10.80	0.40			0.29	9.85	0.47
	0.31	15.90	0.22	10.81	0.72	0.26	17.36	0.33	13.09	0.78
730	0.31	13.19			0.43	0.17	5.81			0.19
			0.17	9.71	0.30			0.22	8.78	0.40
	0.31	18.31	0.16	8.98	0.72	0.23	16.75	0.26	11.74	0.73

This table reports the results of regressing the squared implied volatilities of different maturities on the extracted long- and short-variance factors. We calibrate a two-factor stochastic volatility option pricing model to the weekly data of ATM options each year from 1996 to 2015, using an iterative two-step optimization procedure as in Christoffersen, Heston and Jacobs (2009). The *t*-statistics are adjusted by Newey-West standard errors.

$$\sigma_t^2(\tau) = \alpha(\tau) + \beta_2(\tau)V_{2t} + \varepsilon_t(\tau), \tag{12}$$

and the bivariate regression

$$\sigma_t^2(\tau) = \alpha(\tau) + \beta_1(\tau)V_{1t} + \beta_2(\tau)V_{2t} + \varepsilon_t(\tau), \tag{13}$$

to investigate the relationship between these two extracted variance factors and the implied volatilities. V_{1t} and V_{2t} are the two variance factors in Eq. (9), while $\varepsilon_t(\tau)$ is the residual of the regression for the τ-day implied variance.

Table 10 reports the regression results of the ATM calls and puts. Most implied volatilities are affected by both factors with significant t statistics. Short-maturity implied volatilities are more related to the short-term variance factor, while long-maturity implied volatilities are more related to the long-term variance factor. For example, the long-term variance factor only explains 6% of the variances of the 30-day call's implied volatility, but its explanatory power for the 730-day call's implied volatility increases to 43%. Meanwhile, the adjusted R^2 of the short-term variance factor on 30- and 730-day calls' implied volatilities are 72% and 30%, respectively. Results of ATM puts are similar. Consistent with our hypothesis, long-maturity implied volatilities contain more information about the long-run equilibrium of variance, while short-maturity implied volatility contains more information about short-term variance.

Another interesting finding is that the explanatory power of the two variance factors is higher for short-maturity implied volatilities. For example, the adjusted R^2 of the 30-day call's implied volatility on the two variance factors is 78%, while it is only 72% for the 730-day call's implied volatility. We have similar results for ATM puts. These suggest that the two-factor stochastic volatility model better captures the prices of short-maturity options than long-maturity options.

6 | CONCLUSION

In this paper, we test the out-of-sample predictability of S&P 500 index option implied volatilities. In particular, we evaluate 14 different models that are based on historical implied volatility surface information. We investigate both statistical and economic significance. To examine how long this predictability lasts, we also compare the results at different forecast horizons. We obtain several interesting results.

Using out-of-sample R_{OS}^2 statistics as the statistical measure, we find that several models that use the entire historical implied volatility surface information could predict the implied volatility significantly in the out-of-sample period. These models could forecast the implied volatility up to 1 week ahead for the call options, and up to 20 days ahead for the puts.

Using the gain on Leland's alpha as the economic significance measure, we find that the predictability is of economic significance. The models that use the information of implied volatility surface generate positive gain on Leland's alphas relative to the benchmark model, even after transaction costs are accounted for. During the recent financial crisis, the predictability is weakened but the economic significance becomes stronger. In particular, the VAR(1) model on volatility changes performs well in hedging against the downside risk during the crisis.

By calibrating a two-factor stochastic volatility model to option data, we extract a long-term and a short-term variance component. By regressing different implied volatilities on these two components, we find that short-maturity implied volatilities are more related to the short-term variance factor, while long-maturity volatilities are more related to the long-term variance factor. This helps explain why using them jointly improves the forecast performance.

Our findings have several interesting implications. Our results show the importance of historical implied volatility information up to 1 week. We carry out delta-neutral trading strategies and document the economic significance of option market predictability. The results of significantly positive abnormal returns provide insight of profitable investment opportunities for hedge fund managers, and show an economically effective way of using historical implied volatility curve information for practitioners.

Our results are consistent with Bakshi et al. (1997) and the emerging component volatility models. Both short-term and long-term volatilities should be considered in option pricing models to fully capture the price influence.

ACKNOWLEDGMENTS

We thank Henk Berkman, Jianguo Chen, Rong Chen, Toby Daglish, Wenjin Kang, Xingguo Luo, Ben Marshall, Michael Naylor, Guoshi Tong, Nuttawat Visaltanachoti and seminar participants at Massey University, Renmin University of China, Southwestern University of Finance and Economics, Xiamen University, Zhejiang University, Victoria University of Wellington 2015 Finance

Workshop, 2016 New Zealand Finance Colloquium, the Midwest Finance Association 2016 Annual Meeting, and 2016 FMA European Conference for their helpful comments.Part of this work was done during Guo's visit to the University of Sussex, UK. Guo acknowledges the support of the National Natural Science Fund of China (Grant No. 71503254). Han acknowledges the support of the National Natural Science Foundation of China (NNSFC Grant No. 71471153) and NNSFC Key Project (Grant No. 71631004).

ORCID

Hai Lin http://orcid.org/0000-0003-4709-799X

REFERENCES

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, *52*, 2003–2049.

Bollen, N., & Whaley, R. (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance*, *59*, 711–753.

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

Campbell, J. Y., & Shiller, R. J. (1991). Yield spreads and interest rate movements: A birds eye view. *Review of Economic Studies*, *58*, 495–514.

Cao, J., & Han, B. (2013). Cross-section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, *108*, 231–249.

Cao, J., Han, B., Tong, Q., & Zhan, X. (2017). *Option return predictability* (Working paper). University of Toronto.

Carr, P., & Wu, L. (2007). Stochastic skew in currency options. *Journal of Financial Economics*, *86*, 213–247.

Chalamandaris, G., & Tsekrekos, A. E. (2010). Predictable dynamics in implied volatility surfaces from OTC currency options. *Journal of Banking and Finance*, *34*, 1175–1188.

Chalamandaris, G., & Tsekrekos, A. E. (2011). How important is the term structure in implied volatility surface modeling? Evidence from foreign exchange options. *Journal of International Money and Finance*, *30*, 623–640.

Chen, N., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, *59*, 383–403.

Chiras, D. P., & Manaster, S. (1978). The information content of option prices and a test of market efficiency. *Journal of Financial Economics*, *6*, 213–234.

Christoffersen, P., Heston, S., & Jacobs, K. (2009). The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well? *Management Science*, *55*, 1914–1932.

Christoffersen, P., Jacobs, K., Ornthanalai, C., & Wang, Y. (2008). Option valuation with long-run and short-run volatility components. *Journal of Financial Economics*, *90*, 272–297.

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, *95*, 138–160.

Constantinides, G. M., Jackwerth, J. C., & Savov, A. (2013). The puzzle of index option returns. *Review of Asset Pricing Studies*, *3*, 229–257.

De Fontnouvelle, P., Fisher, R., & Harris, J. (2003). The behavior of bid-ask spreads and volume in options markets during the competition for listings in 1999. *Journal of Finance*, *58*, 2437–2463.

Derman, E., Kani, I., & Zou, J. Z. (1996). The local volatility surface: Unlocking the information in index option prices. *Financial Analysts Journal*, *52*, 25–36.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*, 337–364.

Egloff, D., Leippold, M, & Wu, L. (2010). The term structure of variance swap rates and optimal variance swap investments. *Journal of Financial and Quantitative Analysis*, *45*, 1279–1310.

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *American Economic Review*, *77*, 680–692.

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

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*, 23–49.

Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics*, *5*, 115–146.

Galai, D. (1977). Tests of market efficiency of the Chicago board options exchange. *Journal of Business*, *50*, 167–197.

Gargano, A., Pettenuzzo, D., & Timmermann, A. (2017). Bond return predictability: Economic value and links to the macroeconomy. *Management Science*. https://doi.org/10.1287/mnsc.2017.2829

Goh, J., Jiang, F., Tu, J., & Zhou, G. (2012). *Forecasting bond risk premia using technical indicators* (Working paper). Washington University in St. Louis.

Goncalves, S., & Guidolin, M. (2006). Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business*, *79*, 1591–1635.

Greenwood, R., & Hanson, S. G. (2013). Issuer quality and corporate bond returns. *Review of Financial Studies*, *26*, 1483–1525.

Guo, B., Han, Q., & Zhao, B. (2014). The Nelson-Siegel model of the term structure of option implied volatility and volatility components. *Journal of Futures Markets*, *34*, 788–806.

Han, B. (2007). Investor sentiment and option prices. *Review of Financial Studies*, *21*, 387–414.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, *75*, 1175–1189.

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics*, *146*, 342–350.

Harvey, C. R., & Whaley, R. E. (1992). Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics*, *31*, 43–73.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, *5*, 357–386.

Hogan, S., Jarrow, R., Teo, M., & Warachka, M. (2004). Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *Journal of Financial Economics*, *73*, 525–565.

Johnson, T. L. (2017). Risk premia and the VIX term structure. *Journal of Financial and Quantitative Analysis*, *52*, 2461–2490.

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

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *Journal of Finance*, *68*, 1721–1756.

Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, *186*, 294–316.

Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking and Finance*, *32*, 2401–2411.

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, *44*, 169–203.

Leland, H. E. (1999). Beyond mean-variance: Performance measurement in a nonsymmetrical world. *Financial Analysts Journal*, *55*, 27–36.

Lin, H., Liu, W., Wu, C., & Zhou, G. (2017). *Prediction of Treasury bond returns* (Working paper). Washington University in St. Louis.

Lin, H., Wang, J., & Wu, C. (2014). Predictions of corporate bond excess returns. *Journal of Financial Markets*, *21*, 123–152.

Lin, H., Wu, C., & Zhou, G. (2017). Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. *Management Science*. https://doi.org/10.1287/mnsc.2017.2734

Loh, R. K., & Stulz, R. M. (2014). *Is sell-side research more valuable in bad times?* (Working paper). National Bureau of Economic Research.

Mayhew, S. (2002). Competition, market structure, and bid-ask spreads in stock option markets. *Journal of Finance*, *57*, 931–958.

Muravyev, D., & Pearson, N. D. (2016). *Option trading costs are lower than you think* (Working Paper). Boston College.

Nelson, C. R, & Siegel, A. F (1987). Parsimonious modeling of yield curves. *Journal of Business*, *60*, 473–489.

Neumann, M., & Skiadopoulos, G. (2013). Predictable dynamics in higher-order risk-neutral moments: Evidence from the S&P 500 options. *Journal of Financial and Quantitative Analysis*, *48*, 947–977.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Pettenuzzo, D., Timmermann, A., & Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, *114*, 517–553.

Poon, S.-H., & Pope, P. (2000). Trading volatility spreads: A test of index option market efficiency. *European Financial Management*, *6*, 235–260.

Rapach, D. E., Ringgenberg, M. C., & Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, *121*, 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. *Review of Financial Studies*, *23*, 821–862.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *Journal of Finance*, *68*, 1633–1662.

Sarno, L., Schneider, P., & Wagner, C. (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance*, *37*, 247–267.

Thornton, D. L., & Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Review of Financial Studies*, *25*, 3141–3168.

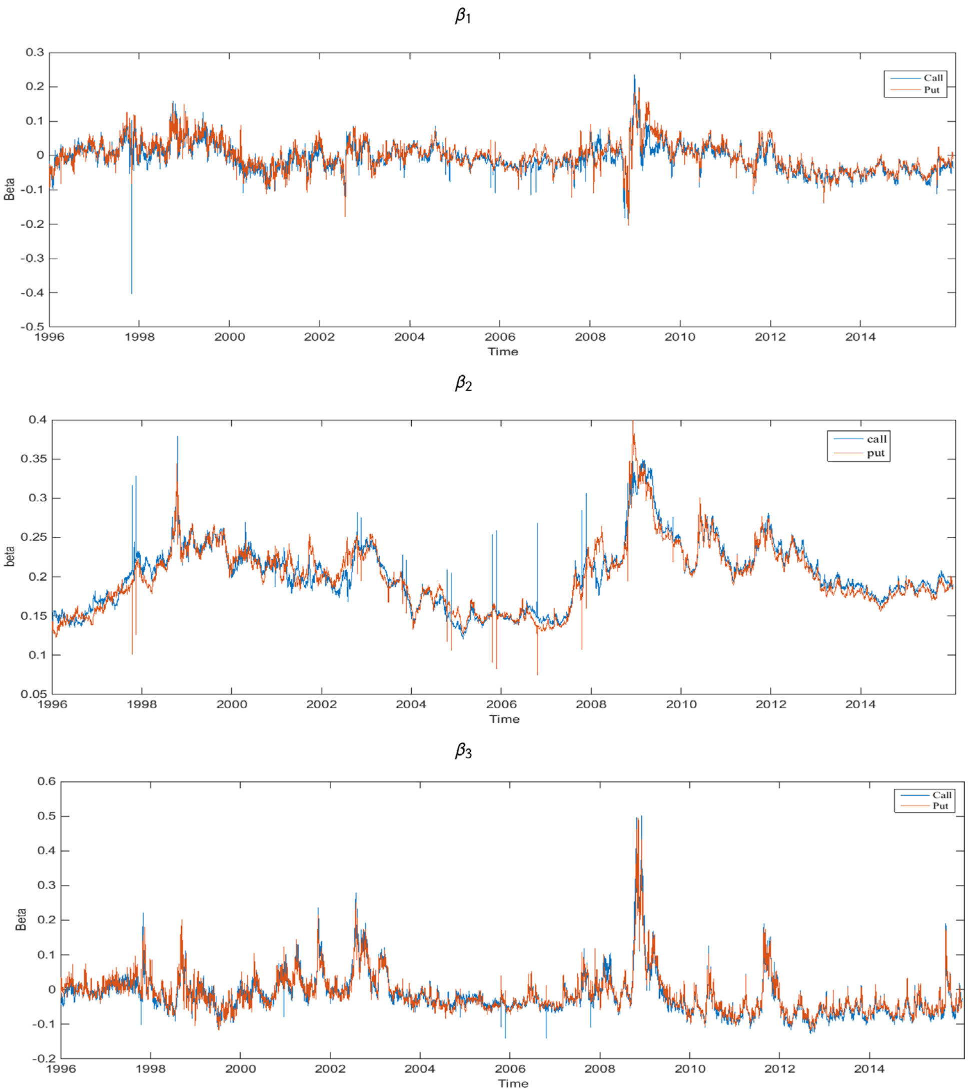
Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, *99*, 216–233.

How to cite this article: Guo B, Han Q, Lin H. Are there gains from using information over the surface of implied volatilities? *J Futures Markets*. 2018;38:645–672. https://doi.org/10.1002/fut.21903

APPENDIX 1

Coefficients of NS model

This graph plots the time series of β_{1t}, β_{2t}, and β_{3t} that are calibrated from the ATM implied volatility curve information using the NS model [Color figure can be viewed at wileyonlinelibrary.com]



APPENDIX 2

Calibration results of two-factor stochastic volatility option pricing model

This appendix reports the calibration results of the two-factor stochastic volatility option pricing model of Christoffersen et al. (2009).

Year	Call								Put							
	a_1/b_1	b_1	a_2/b_2	b_2	σ_1	$\sigma_2(\times 100$	ρ_1	ρ_2	a_1/b_1	b_1	a_2/b_2	b_2	σ_1	$\sigma_2(\times 100$	ρ_1	ρ_2
1996	0.08	0.25	0.03	1.07	0.02	0.09	−0.82	−0.60	0.07	0.00	0.10	0.74	0.03	0.40	−0.87	−0.69
1997	0.04	0.24	0.06	1.82	0.02	0.10	−0.83	−0.66	0.10	0.24	0.08	1.18	0.02	0.37	−0.84	−0.67
1998	0.00	0.25	0.08	0.88	0.03	0.05	−0.81	−0.65	0.05	0.24	0.10	1.67	0.04	0.06	−0.84	−0.67
1999	0.00	0.25	0.07	2.73	0.04	0.07	−0.90	−0.68	0.08	0.18	0.08	3.63	0.05	0.58	−0.83	−0.68
2000	0.10	0.24	0.03	1.08	0.02	0.07	−0.83	−0.67	0.10	0.13	0.08	0.94	0.02	0.05	−0.83	−0.67
2001	0.08	0.22	0.03	3.59	0.01	0.05	−0.80	−0.60	0.10	0.11	0.07	0.97	0.02	0.05	−0.87	−0.69
2002	0.08	0.2	0.05	3.97	0.00	0.05	−0.79	−0.65	0.10	0.18	0.07	1.52	0.02	0.05	−0.86	−0.73
2003	0.00	0.07	0.04	3.71	0.01	0.05	−0.93	−0.70	0.09	0.06	0.07	1.43	0.03	0.10	−0.84	−0.76
2004	0.00	0.25	0.06	0.88	0.03	0.05	−0.83	−0.71	0.02	0.20	0.08	2.16	0.04	0.05	−0.85	−0.68
2005	0.00	0.25	0.07	9.99	0.11	0.06	−0.71	−0.50	0.03	0.08	0.07	1.75	0.04	0.05	−0.81	−0.70
2006	0.04	0.15	0.03	3.32	0.03	0.05	−0.84	−0.70	0.10	0.14	0.08	1.62	0.04	0.64	−0.86	−0.69
2007	0.02	0.05	0.04	4.94	0.02	0.35	−0.83	−0.73	0.10	0.10	0.08	1.21	0.03	0.05	−0.81	−0.71
2008	0.00	0.17	0.07	6.54	0.00	0.07	−0.79	−0.56	0.10	0.25	0.11	1.33	0.04	0.05	−0.83	−0.72
2009	0.00	0.25	0.10	0.96	0.03	0.05	−0.82	−0.72	0.01	0.19	0.14	1.73	0.05	0.08	−0.82	−0.73
2010	0.00	0.25	0.08	3.35	0.07	0.07	−0.78	−0.56	0.07	0.00	0.10	3.22	0.07	0.41	−0.86	−0.71
2011	0.00	0.22	0.08	2.67	0.04	0.05	−0.84	−0.65	0.10	0.03	0.11	2.23	0.06	0.08	−0.84	−0.61
2012	0.00	0.25	0.08	3.87	0.03	46.26	−0.83	−0.83	0.07	0.00	0.14	2.09	0.07	12.88	−0.90	−0.55
2013	0.10	0.19	0.04	2.65	0.05	0.09	−0.85	−0.73	0.06	0.00	0.12	1.26	0.05	0.62	−0.89	−0.68
2014	0.09	0.25	0.04	4.69	0.05	2.97	−0.83	−0.71	0.10	0.18	0.09	1.85	0.05	0.78	−0.85	−0.70
2015	0.02	0.25	0.07	1.38	0.04	0.05	−0.79	−0.73	0.10	0.24	0.10	1.53	0.04	0.43	−0.85	−0.69