Forecasting the Term Structure of Implied Volatilities

Biao Guo

Renmin University of China

Qian Han

Xiamen University

Hai Lin

Victoria University of Wellington*

First draft: September 13, 2013

Current version: October 9, 2015

*Correspondence: Hai Lin, School of Economics and Finance, Victoria University of Wellington, Wellington 6140, New Zealand. Phone: (64)-4-463-5239 and email: hai.lin@vuw.ac.nz. We thank Jianguo Chen, Rong Chen, Wenjin Kang, Xingguo Luo, Ben Marshall, Michael Naylor and seminar participants at Massey University, Renmin University of China, Southwestern University of Finance and Economics, Xiamen University and Zhejiang University for helpful comments. Han acknowledges the support from the National Natural Science Fund of China (Grant No.71471153). Lin acknowledges the support from the National Natural Science Fund of China (Grant No. 71301137, 71371161, 71101121) and the Humanities and Social Science Fund of Chinese Ministry of Education (Grant No. 11JYC790014).

Forecasting the Term Structure of Implied Volatilities

Neumann and Skiadopoulos (2013) document that although the implied volatilities are predictable, their economic profits become insignificant once the cost is accounted for. We show that the trading strategies based on the predictability of implied volatilities could generate significant risk-adjusted returns after controlling for the transaction cost. The implied volatility curve information is useful for the out-of-sample forecast of implied volatilities up to one week. Short-maturity implied volatilities tend to be more predictable than long-maturity implied volatilities. Although the long-maturity options are much less traded than the short-maturity options, their implied volatilities provide much more information on the price discovery.

I. Introduction

There is a large body of literature studying the predictability of implied volatility that serves as a key measure of option price. Most studies focus on two research questions: First, are the implied volatilities predictable? Second, is there significant economic profit from such predictability? For the first question, the literature generally documents that the implied volatilities are predictable, such as Harvey and Whaley (1992), Gonclaves and Guidolin (2006), Konstantinidi, Skiadopoulos and Tzagkaraki (2008), Chalamandaris and Tsekrekos (2010, 2011) and Neumann and Skiadopoulos (2013). On the other hand, the results of economic significance are quite mixed. Harvey and Whaley (1992), Gonclaves and Guidolin (2006), Konstantinidi, Skiadopoulos and Tzagkaraki (2008), Chalamandaris and Tsekrekos (2010, 2011) and Neumann and Skiadopoulos (2013) show that although the implied volatilities are predictable, their economic profits become insignificant once the cost is accounted for. However, Galai (1977), Chiras and Manaster (1978), Poon and Pope (2000) and Hogan et al. (2004) find significant excess returns of trading strategies even if the transaction cost is considered.[1]

Several interesting questions arise from these studies and are still open for discussion. For example, what kind of information is useful for the prediction? How long does the predictability last? Is the predictability robust over time and over different option series? In this paper, we conduct a comprehensive study on the predictability of implied volatilities by focusing on six major questions.

The first is whether or not the historical information of implied volatility curve is useful for the forecast of implied volatilities. We address this issue by conducting a horse race of 16 models and compare their out-of-sample performance with the random walk model as a

[1]Other relevant studies include Klemkosky and Resnick (1979), Philips and Smith (1980) and Bodurtha and Courtadon (1986). They carry out put-call parity boundary tests and find that the put-call parity holds in general.

benchmark. These models include two adapted Nelson and Siegel models used by Diebold and Li (2006) for Treasury securities and by Chalamandaris and Tsekrekos (2011) for currency options, eight time series models similar to Diebold and Li (2006), five combination models as in Rapach, Strauss and Zhou (2010) and Mallows Model Averaging (MMA) combination as in Hansen (2007, 2008). We find that historical information plays a role in the prediction of implied volatilities during a short period of time. For example, when we use the daily data to forecast the one-month implied volatility one day ahead and five days ahead, the best out-of-sample R^2 is 5.34% and 8.42%, respectively. Both of them are statistically significant.

The second question is how persistent the predictability of implied volatility is. We find that when the daily data are used, implied volatilities are highly predictable up to five days ahead. In general, these models lost their predictive power beyond a week, suggesting that only the historical information within one week is important for the forecast of the index option market.

The third question is what historical information is useful in the forecast. The historical information not only includes the time series information of one particular maturity, but also the cross-sectional information of other maturities. Bakshi, Cao and Chen (2000) show that long-dated options have information not readily available from short-dated options. This raises the question of whether incorporating the whole or part of the term structure of implied volatilities can help the prediction. Our results reveal that the models that use the information of the whole implied volatility curve perform much better than the models that only use the historical time series information of a single maturity. The average out-of-sample R^2 of one-day forecasts and five-day forecasts across maturities for the at-the-money (ATM) call option could reach as high as 4.96% and 6.39%, respectively, for the models using the whole curve information, compared with the 0.24% and 1.15%, respectively, for the models that do not use it. This suggests that when we forecast the implied volatility of a particular maturity, it is helpful to consider the historical information of other maturities.

Consistent with Rapach, Strauss and Zhou (2010), the combination of individual forecasts also provides stable and significant results.

The fourth question is whether the predictability has economic value. We construct a trading strategy based on the forecast by each of these different models, and compare the trading portfolio performance with the benchmark model. Using the Sharpe ratio and Leland's Alpha as the performance measure, we find that these models generate significant economic profits up to five days even after moderate transaction costs are considered. For example, when the daily data are used, the trading strategy based on the mean combination forecast of one day ahead generates a Sharpe ratio of 0.31 and a Leland's Alpha of 12.38%. Both are significant at the 1% level. Similarly, the trading strategy based on the mean combination forecast of five days ahead generates a Sharpe ratio of 0.10 and a Leland's Alpha of 5.07%. This finding is different from most other literature that found no predictability of the option market after considering the transaction cost.

The fifth question is whether the predictability is robust over time and over different option series. To address the first part, we conduct a sub-sample analysis using the data during the recent 2007-2009 financial crisis period. We find that the predictability becomes slightly weaker. However, implied volatilities can still be predicted one day ahead. Moreover, their economic significance becomes stronger. We then use the ATM put option data and the call option data with different moneyness to test the robustness of our findings across different option series, and the main results still hold.

The sixth question is, since the implied volatilities can be better predicted by using the whole implied volatility curve, which maturities contain more information for the price discovery of term structure of implied volatilities? We use the Hasbrouck (1995) and Gonzalo-Granger (1995) price discovery measures to assess the information share of each maturity option, and link them with the trading statistics. We find that although the long-maturity options are much less traded, their implied volatilities contain more information of price discovery than those of short-maturity options.

Our study contributes to the literature in several ways. First, our findings shed light on volatility modelling, portfolio management and policy implementation. We evaluate an extensive set of 16 models in the option market, including the models that use the time series information only and the models that use both times series and cross-sectional information. Such tests are important to answer the question of which information is useful in the forecast and what models should be used to capture the dynamics of implied volatility. Our finding that the whole implied volatility term structure plays roles in forecasting, in particular that long-maturity implied volatilities provide more useful information than short-maturity implied volatilities, implies that a one factor model is not enough for volatility modelling. In this regard, we provide empirical evidence to the emerging component volatility models (see, for example, Christoffersen et al., 2008; Christoffersen, Heston and Jacobs, 2009). Our work also extends Rapach, Strauss and Zhou (2010)'s out-of-sample tests of predictability on stock market to option market.

Second, this paper documents that more trading (and possibly better liquidity) does not necessarily mean faster speed of information adjustment, and contributes to the debate of the relationship between liquidity and price discovery. The relationship between price discovery and liquidity is still a controversy in the literature. The common hypothesis is that price discovery is positively related with liquidity. For example, Kwan (1996) finds that stock returns predict future bond yield changes because the stock market is more liquid. Similarly, Chakravarty, Gulen and Mayhew (2004) show that on option market although the illiquid out-of-money (OTM) options have higher price discovery measures than liquid in-the-money (ITM) options, the price discovery is still positively related to liquidity measures such as trading volume and price spread once the leverage is controlled. However, there are other findings that do not support this hypothesis. Hotchkiss and Ronen (2002) find that although corporate bond market is much less liquid, corporate bond returns cannot be predicted by past stock returns based on a sample of 20 high-yield bonds from the National Association of

Securities Dealers (NASD). Barclay and Hendershott (2003) study the effect of trading after hours on the price discovery over the 24-hour day and find that it is possible to generate significant price discovery with very little, but very informative, trading. Zhu (2014) shows that price discovery can improve along with reduced liquidity. What is crucial to price discovery is informed traders, not necessarily liquidity. If the trading has much noise and is subject to behavior bias, more trading might result in the deviation from its equilibrium level, and provide a noisy information signal. On the other hand, if the trading is mainly from informed trading, it will only take few transactions to fully reflect the impact of information shock. Therefore, thinly traded long-maturity options may contain more information than heavily traded short-maturity options and they are critical for price discovery.

This paper is most closely related to Gonclaves and Guidolin (2006) and Neumann and Skiadopoulos (2013) in that we all test the out-of-sample predictability of S&P 500 index implied volatilities both statistically and economically. Gonclaves and Guidolin (2006) propose a two-stage method to predict the dynamics of the S&P 500 index implied volatility surface (IVS) and find that the IVS is predictable one day ahead. However, the trading strategies based on this predictability generate no significant profits once the transaction cost is included. Neumann and Skiadopoulos (2013) model the evolution of IVS from the perspective of higher-order moments and conclude that they are highly predictable over different forecasting horizons, even on a weekly and monthly basis, but the economic significance disappears after considering the transaction costs.

Our finding that the cross-sectional information of other maturities is useful for prediction supports their modelling on the IVS. However, we find a significant economic profit even if the transaction costs are considered, which is different from theirs. Several other differences also separate our paper from theirs. First, we test a larger set of models. Both Gonclaves and Guidolin (2006) and Neumann and Skiadopoulos (2013) use five or fewer models, and hence their results may be less convincing. Second, analogous to Diebold and Li (2006), we choose all forecasting models as either AR(1) or VAR(1) or their combinations, compared with the sophisticated models in Gonclaves and Guidolin (2006) and Neumann and Skiadopoulos (2013). Rather than predicting the implied volatility directly, they predict the IVS model parameters using a time series model, and then put the predicted parameters into the IVS model and obtain the forecasted IVS. These two-stage predictions are subject to high prediction errors. Error-in-variable (EIV) problems may arise when a model is first fitted to the IVS and then used in the forecast.[2] Third, we examine the impacts of the financial crisis on the predictability of implied volatilities. Gonclaves and Guidolin (2006) use data only from 1992 to 1996; the sample period from 1996 to 2010 in Neumann and Skiadopoulos (2013) covers the crisis, but the authors do not conduct a sub-sample analysis. Our finding that the predictability has even stronger economic values during the crisis corroborates our conclusion that implied volatilities can be predicted, and be profitable even after transaction costs are accounted for.

The rest of the paper is structured as follows. Section 2 introduces our empirical methodologies, including the 16 models to be tested, the out-of-sample performance evaluation criterion and the price discovery measures. Section 3 discusses the data and presents major empirical results. Section 4 provides the robustness checks, including the sub-sample analysis using the data covering the recent crisis period and the out-of-sample performance of other option series. Section 5 reports the results of price discovery analysis. Section 6 concludes the paper.

II. Empirical Methodology

This section outlines the models to be evaluated, the statistical and economic significance measures for evaluating the prediction performances and the price discovery measures. Our

[2]As shown later, the models that use the Nelson-Siegel (NS) model to fit the term structure of implied volatility and then forecast them by forecasting the NS model parameters perform worse than the other models that predict the implied volatility directly.

study focuses on the out-of-sample test. Suppose we have the 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 could compare the forecasted implied volatility and realized implied volatility to calculate the out-of-sample forecast errors. Different prediction models will have different forecast errors. This procedure is repeated from time m to $T-h$.

A. Prediction model

The Nelson and Siegel (1987) model and its extension (Diebold and 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 Derman, Kani and Zou, 1996; Christoffersen, Heston and Jacobs, 2009). 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 volatility term structures exhibit a high degree of time variation and cross-sectional variation. Since the Nelson-Siegel model is an empirical model, it can be borrowed directly to model the term structure of implied volatility. We fit the implied volatility curve $\sigma_t(\tau)$ using the Nelson-Siegel model,

$$\sigma_t(\tau) = \beta_{1t} + \beta_{2t}\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} + \beta_{3t}(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)), \quad (1)$$

where τ is time to maturity and the parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ are estimated by ordinary least squares (OLS) with λ_t fixed at a pre-specified value of 0.0147.[3] The loading on β_{1t} is 1, a

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

constant that does not decay to zero in the limit; hence β_{1t} may be viewed as a long-term factor. The loading on β_{2t} 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} is $\frac{1-\exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau)$, which starts at 0 and increases, and then decays to zero, hence it may be viewed as a medium-term factor.

Besides the Nelson-Siegel model, we consider eight time series models following Diebold and Li (2006), five combination models as in Rapach, Strauss and Zhou (2010) and the Mallows Model Averaging (MMA) combination as in Hansen (2007, 2008). Table 1 lists all the 16 models to be evaluated in this paper and the benchmark model used for comparison. We forecast the implied volatility curve h days ahead with the following models:

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

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

where $\hat{\beta}_{i,t+h} = c_{0,i} + c_{1,i}\hat{\beta}_{i,t}, i = 1, 2, 3.$

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

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

where $\hat{\beta}_{t+h} = c_0 + c_1\hat{\beta}_t; \hat{\beta}_t = \begin{bmatrix} \hat{\beta}_{1,t} & \hat{\beta}_{2,t} & \hat{\beta}_{3,t} \end{bmatrix}^T.$

(3) Slope regression: $\hat{\sigma}_{t+h}(\tau) - \sigma_t(\tau) = c_0(\tau) + c_1(\tau)(\sigma_t(\tau) - \sigma_t(30)).$

(4) AR(1) on volatility levels: $\hat{\sigma}_{t+h}(\tau) = c_0(\tau) + c_1(\tau)\sigma_t(\tau).$

(5) VAR(1) on volatility levels: $\hat{\sigma}_{\mathbf{t+h}} = c_0 + c_1 \sigma_{\mathbf{t}}$, where $\sigma_{\mathbf{t}} = \begin{bmatrix} \sigma_t(30) \\ \sigma_t(91) \\ \sigma_t(152) \\ \sigma_t(365) \\ \sigma_t(730) \end{bmatrix}$.

(6) VAR(1) on volatility changes: $\hat{\mathbf{z}}_{\mathbf{t+h}} = c_0 + c_1 \mathbf{z}_{\mathbf{t}}$, where $\mathbf{z}_{\mathbf{t}} = \begin{bmatrix} \sigma_t(30) - \sigma_{t-h}(30) \\ \sigma_t(91) - \sigma_{t-h}(91) \\ \sigma_t(152) - \sigma_{t-h}(152) \\ \sigma_t(365) - \sigma_{t-h}(365) \\ \sigma_t(730) - \sigma_{t-h}(730) \end{bmatrix}$.

(7) ECM(1) with one common trend: $\hat{\mathbf{z}}_{\mathbf{t+h}} = c_0 + c_1 \mathbf{z}_{\mathbf{t}}$, where $\mathbf{z}_{\mathbf{t}} = \begin{bmatrix} \sigma_t(30) - \sigma_{t-h}(30) \\ \sigma_t(91) - \sigma_t(30) \\ \sigma_t(152) - \sigma_t(30) \\ \sigma_t(365) - \sigma_t(30) \\ \sigma_t(730) - \sigma_t(30) \end{bmatrix}$.

(8) ECM(1) with two common trends: $\hat{\mathbf{z}}_{\mathbf{t+h}} = c_0 + c_1 \mathbf{z}_{\mathbf{t}}$, where $\mathbf{z}_{\mathbf{t}} = \begin{bmatrix} \sigma_t(30) - \sigma_{t-h}(30) \\ \sigma_t(91) - \sigma_{t-h}(91) \\ \sigma_t(152) - \sigma_t(30) \\ \sigma_t(365) - \sigma_t(30) \\ \sigma_t(730) - \sigma_t(30) \end{bmatrix}$.

(9) AR(1) regression on three principal components. We first conduct a principal component analysis on the 10 volatilities time series data. Denote the largest three eigenvalues by λ_1, λ_2 and λ_3, with associated eigenvectors q_1, q_2 and q_3, and the first three principal components $\mathbf{x_t} = \begin{bmatrix} x_{1t} & x_{2t} & x_{3t} \end{bmatrix}^T$. We first forecast $\mathbf{x_{t+1}}$ with a univariate AR(1) model, $\hat{x}_{i,t+h} = c_{0,i} + c_{1,i} x_{i,t}, i = 1, 2, 3$, and then generate forecasts for volatilities as $\hat{\sigma}_{t+h}(\tau) = q_1(\tau)\hat{x}_{1,t+h} + q_2(\tau)\hat{x}_{2,t+h} + q_3(\tau)\hat{x}_{3,t+h}$.

(10) VAR(1) on empirical level, slope and curvature: $\hat{\sigma}_{\mathbf{t+h}} = c_0 + c_1\mathbf{F_t}$, where $\mathbf{F_t} = \begin{bmatrix} \sigma_t(365) \\ \sigma_t(365) - \sigma_t(30) \\ 2\sigma_t(122) - (\sigma_t(365) + \sigma_t(30)) \end{bmatrix}$. 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 of 122-day implied volatility minus the sum of the 365-day and 30-day implied volatilities.

Researchers have shown that combination forecasts typically outperform individual forecasts both statistically and economically. For example, Rapach, Strauss and Zhou (2010) find that combination delivers consistent forecast gains for equity premium predictions. So besides the above forecasts in (1) $\sim$ (10), we further combine them as $\hat{\sigma}_{c,t+h}(\tau) = \sum_{k=1}^{10} w_{k,t}(h,\tau)\hat{\sigma}_{k,t+h}(\tau)$, which yields five forecasts depending on the combination weight $w_{k,t}(h,\tau)$:

(11) The mean combination forecast: $w_{k,t}(h,\tau) = 1/10$.

(12) The median combination forecast: the median of $\hat{\sigma}_{k,t+h}(\tau), k = 1 \sim 10$.

(13) The trimmed mean combination forecast: $w_{k,t}(h,\tau) = 0$ for the smallest and largest forecasts and $w_{k,t}(h,\tau) = 1/8$ for the remaining forecasts.

(14) DMSPE (discount mean square prediction error) combination forecast one: $w_{k,t}(h,\tau) = \frac{\phi_{k,t}^{-1}(h,\tau)}{\sum_{i=1}^{10}\phi_{i,t}^{-1}(h,\tau)}$, where $\phi_{k,t}(h,\tau) = \sum_{j=m}^{t-h}\theta^{t-h-j}(\sigma_{j+h}(\tau) - \hat{\sigma}_{k,j+h}(\tau))^2$, θ is a discounting factor deciding the size of weights given to the recent forecasts, m is the starting time of out-of-sample forecast. We take $\theta = 1$ for no discounting to remote forecast.

(15) DMSPE (discount mean square prediction error) combination forecast two: Same as (14) except that we take $\theta = 0.9$ to give greater weight to recent forecast.

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. We also consider it in our forecast.

(16) MMA combination. Let $\mathbf{w_t}(\mathbf{h},\boldsymbol{\tau}) = [w_{1,t}(h,\tau)...w_{10,t}(h,\tau)]^T$ be the weight vector of the individual forecast, $\hat{\sigma}_{\mathbf{j+h}}(\boldsymbol{\tau}) = [\hat{\sigma}_{1,j+h}(\tau)...\hat{\sigma}_{10,j+h}(\tau)]^T$be the vector of individual h-day forecast of τ-day implied volatility at time j, and $\mathbf{G} = [g(1)...g(10)]^T$ be the vector of predictor number used in the individual forecast. MMA combination forecast set $\mathbf{w_t}(\mathbf{h},\boldsymbol{\tau})$ to minimize $C_t(h,\tau)$ with the conditions that all $w_{k,t}(h,\tau)$ are non-negative and $\sum_{k=1}^{10} w_{k,t}(h,\tau) = 1$. $C_t(h,\tau)$ is calculated by $C_t(h,\tau) = \sum_{j=m}^{t-h}(\sigma_{j+h}(\tau) - \hat{\sigma}_{\mathbf{j+h}}(\boldsymbol{\tau})^T \mathbf{w_t}(\mathbf{h},\boldsymbol{\tau}))^2 + 2\mathbf{w_t}(\mathbf{h},\boldsymbol{\tau})^T \mathbf{G} s^2$, where s^2 is an estimate of the variance of residuals from the largest fitted model.

The benchmark model is a random walk model, which assumes that the historical information is not useful and uses the current value to predict the implied volatility, i.e., $\hat{\sigma}_{t+h}(\tau) = \sigma_t(\tau)$.

[Insert Table 1 here]

B. Out-of-sample forecast evaluation

In order to check the efficiency of the prediction models, we calculate the out-of-sample R^2 statistics of each model for each maturity, given by

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

For model (14) and model (15) that require the hold-out period (p) in calculating the optimal weight, the forecasting errors used to calculate the $R^2_{OS}(\tau)$ start since $m + p$ until $T - h$. $\hat{\sigma}(\tau)$ and $\bar{\sigma}(\tau)$ is the forecast of implied volatility by model (1) to model (16) and the forecast by the benchmark random walk model, respectively. A positive $R^2_{OS}(\tau)$ indicates that the prediction model outperforms the benchmark model. We calculate the

MSPE-adjusted statistic to test the significance of $R^2_{OS}(\tau)$. Define

$$f_{t+h}(\tau) = [(\sigma_{t+h}(\tau) - \bar{\sigma}_{t+h}(\tau))]^2 - [(\sigma_{t+h}(\tau) - \hat{\sigma}_{t+h}(\tau))^2 - (\hat{\sigma}_{t+h}(\tau) - \bar{\sigma}_{t+h}(\tau))^2], \quad (5)$$

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 $R^2_{OS}(\tau)$. We use Hodrick (1992) to calculate the standard errors that are robust for the data overlapping. In order to test the overall efficiency of prediction models, we also calculate the overall out-of-sample R^2 statistics of each model by

$$R^2_{OS} = 1 - \frac{\sum_\tau \sum_{j=m}^{T-h} (\sigma_{j+h}(\tau) - \hat{\sigma}_{j+h}(\tau))^2}{\sum_\tau \sum_{j=m}^{T-h} (\sigma_{j+h}(\tau) - \bar{\sigma}_{j+h}(\tau))^2}, \quad (6)$$

and test its significance by

$$f_{t+h} = \sum_\tau [(\sigma_{t+h}(\tau) - \bar{\sigma}_{t+h}(\tau))]^2 - \sum_\tau [(\sigma_{t+h}(\tau) - \hat{\sigma}_{t+h}(\tau))^2 - \sum_\tau [(\hat{\sigma}_{t+h}(\tau) - \bar{\sigma}_{t+h}(\tau))]^2. \quad (7)$$

C. Economic significance

We follow Goncalves and Guidolin (2006) and Neumann and Skiadopoulos (2013) to evaluate the trading performance of out-of-sample forecasts and test whether the models can generate abnormal profits. The trading strategies are simply based on the forecasted volatility. Specifically, at date t we long (short) an option if the forecasted volatility for that maturity at date $t + h$ is larger (smaller) than the current volatility. Consistent with previous sections, we consider options with maturity 30, 91, 152, 365 and 730 days; thus, on each date our portfolio includes five option contracts with trading status either long, short or no trading. We hold the portfolio the same time as our forecasting horizon and repeat the trading in the out-of-sample period. Analogous to Goncalves and Guidolin (2006), we delta-hedge our option position by buying (selling) Δ shares of the S&P 500 index if we short (long) the options. The hedge ratio is calculated using the Black-Scholes option pricing formula with the forecasted volatility. We invest \$1000 on each trading date in our portfolio. Let P_t be the value of a unit portfolio at date t,

$$P_t = \sum_{i \in long} (C_{it} - S_t \Delta_{it}) - \sum_{i \in short} (C_{it} - S_t \Delta_{it}), \tag{8}$$

where C is the call option price, S is the S&P 500 index price and long and short represents the option we either buy or sell.[4] Since we invest \$1000 in the portfolio, the total units are $Q_t = |\frac{1000}{P_t}|$. The change of portfolio value at date $t+h$ becomes

$$\begin{aligned} \Delta V_{t+h} = Q_t \sum_{i \in long} (C_{i,t+h} - C_{it}) + Q_t \sum_{i \in short} (C_{it} - C_{i,t+h}) \\ -Q_t(S_{t+h} - S_t) \sum_{i \in long} \Delta_{it} + Q_t(S_{t+h} - S_t) \sum_{i \in short} \Delta_{it} \end{aligned} \tag{9}$$

Finally, we calculate the return of our portfolio as $\Delta V_{t+h}/1000$. We calculate the Sharpe ratio and Leland's (1999) Alpha to gauge the economic performance of these trading strategies, together with the Politis and Romano (1994) stationary bootstrap method to test the significance of these measures. The Sharpe ratio is calculated as $\frac{E(r_a - r_f)}{\sqrt{\text{var}(r_a - r_f)}}$, where r_a is the portfolio return, r_f is the risk-free rate approximated by the U.S. LIBOR rate. Leland's Alpha takes into account the deviation of portfolio returns from normal distribution and equals to $\alpha_p = E[r_a] - \beta_p(E[r_m - r_f]) - r_f$, where r_m denotes the market return approximated by the S&P 500 index return, $\beta_p = \frac{\text{cov}(r_a, -(1+r_m)^{-\gamma})}{\text{cov}(r_m, -(1+r_m)^{-\gamma})}$ measures systematic risk and $\gamma = \frac{\ln(E[1+r_m]) - \ln(1+r_f)}{\text{var}(\ln(1+r_m))}$ measures the relative risk aversion. Since our benchmark model is a random walk model and the forecasted volatility at $t+h$ will always equal to the current volatility, no trading signal is generated for the whole sample. The economic measures of portfolio return using the random walk model in the forecast will always be zero. Therefore, these measures also indicate the difference of portfolio performance between the tested

[4]The option prices are calculated using Black-Scholes option pricing formula.

model and the benchmark model. A larger-than-zero Leland's Alpha thus indicates the trading strategy generates an excess return over the benchmark model. Both the Sharpe ratio and Leland's Alpha are finally annualized.

D. Price discovery

Even if there is an equilibrium relationship between the implied volatilities of different maturities, they may deviate from each other in the short run due to different speeds of information adjustment. Questions arise as to which maturity options respond more quickly to the information shocks, and therefore their implied volatilities reveal more information for the future implied volatility. We employ two price discovery measures suggested by Gonzalo and Granger (1995) and Hasbrouck (1995) to assess the information contents of different maturity options. In brief, the Gonzalo-Granger method decomposes the implied volatility into permanent and transitory components, and links the permanent component with the long-term level. It depends on the adjustment speed of the error-correction term or how the implied volatilities in each maturity change in response to the disequilibrium in the previous period. On the other hand, the Hasbrouck measure not only considers the speed of adjustment, but also incorporates the innovations of two implied volatilities through their covariance structure.

We first estimate the vector error-correction model (VECM) of implied volatilities with two different maturities. Denote $Y_t = [\sigma_t(\tau_1)\ \sigma_t(\tau_2)]^T$, the VECM model can be written as

$$\Delta Y_t = c_0 + \delta\omega^T Y_{t-1} + \sum_{j=1}^{j=q} \Psi_i \Delta Y_{t-1} + \epsilon_t, \tag{10}$$

where c_0 is the constant term, $\omega = [1\ \eta]^T$ is the cointegration vector with the first element normalized to one and $\delta = [\delta_1\ \delta_2]^T$ is the response vector for the error-correction term. ϵ_t are serially uncorrelated innovations with mean zero and a covariance matrix $Var(\epsilon_t)$ with diagonal elements σ_1^2 and σ_2^2 and off-diagonal elements $\rho\sigma_1\sigma_2$. We first test the cointegration relationship between the two pairs. If they are cointegrated, we then use the parameters from the VECM model to construct the price discovery measures.

The Gonzalo-Granger (G) measures for $\sigma_t(\tau_1)$ and $\sigma_t(\tau_2)$ are calculated as

$$G_1 = \frac{-\delta_2}{\delta_1 - \delta_2}, G_2 = \frac{\delta_1}{\delta_1 - \delta_2}. \tag{11}$$

The Hasbrouck measure (H) is defined as

$$H_1(u) = \frac{(-\delta_2\sigma_1 + \delta_1\rho\sigma_2)^2}{(-\delta_2\sigma_1 + \delta_1\rho\sigma_2)^2 + (\delta_1\rho_2\sqrt{1-\rho^2})^2}, \tag{12}$$

$$H_2(l) = \frac{(\delta_1\rho_2\sqrt{1-\rho^2})^2}{(-\delta_2\sigma_1 + \delta_1\rho\sigma_2)^2 + (\delta_1\rho_2\sqrt{1-\rho^2})^2} \tag{13}$$

where u indicates the upper bond and l indicates the lower bond. Reversing the order in the vector of volatility series gives the upper bound $H_2(u)$ and the lower bound $H_1(l)$. We use the average of these two bounds as the Hasbrouck measure of price discovery.

Using the individual price discovery measures calculated from the VECM models run against other maturity implied volatilities, we also calculate the mean price discovery measure. For example, we run the VECM models for the 30-day implied volatility with 90-day, 152-day, 365-day and 730-day implied volatilities, respectively, and obtain four individual price discovery measures of 30-day implied volatility. The mean of those four individual measures is used to assess the overall price discovery function of the 30-day option for the term structure of implied volatilities.

III. Data and empirical results

Our sample includes the daily data of implied volatilities of S&P 500 index options from 1996 to 2011. We use the volatility surfaces taken from the Ivy DB OptionMetrics database, with 10 different time-to-maturities (30, 60, 91, 122, 152, 182, 273, 365, 547 and 730 days) on each observation date. Since not all time-to-maturities are traded on each date, OptionMetrics interpolates the surface to obtain the missing data. We select the ATM call options, as they are the most liquidly traded in the market.[5]

Table 2 reports the mean, maximum, minimum, standard deviation and autocorrelation of implied volatilities with different maturities. The volatility curve is upward sloping, and long-maturity implied volatility has a smaller standard deviation than short-maturity implied volatility. For example, the 730-day implied volatility has a mean of 20.45% and a standard deviation of 4.78%, while the 30-day implied volatility has a mean of 20.07% and a standard deviation of 7.81%. The different persistence across maturities indicates the necessity of modelling the long- and short-maturity implied volatilities separately.

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

[Insert Table 2 here]

[Insert Figure 1 here]

We fit the implied volatility curve using the Nelson-Siegel model by OLS on each obser-

[5] We also extend our studies to the ATM put options and the call options with other moneyness in the robustness check.

vation date. The 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, which 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, together with the close movement between β_{2t} and the empirical slope lines, indicating that β_{2t} reflects the short-term volatility component and can be interpreted as a slope factor. Our findings so far provide strong empirical support for the decomposition of volatility into long- and short-term volatility with a Nelson-Siegel model. [6]

A. Out-of-sample forecast results

We start the out-of-sample forecast in 2002. We estimate parameters using a recursive (expanding) window. We forecast the implied volatility one day, five days and 20 days ahead. The holdout out-of-sample period for model (14) and model (15) is set as 60 days.

Table 3 reports the out-of-sample R^2_{OS} statistics for all the models. The up, middle and down panels report the results of the forecast one day, five days and 20 days ahead, respectively. A positive R^2_{OS} suggests the model outperforms the benchmark random walk model. The statistical significance is assessed with MSPE-adjusted statistics and Hodrick's (1992) standard errors for all positive values.

The results in the up panel show that most of our models are able to beat the benchmark on a one day forecast. For example, 13 out of 16 models generate positive R^2_{OS} 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} and are significant at the 1% level. The advantages of models over the benchmark deteriorate for long-maturity implied volatilities. Still, there are eight models that outperform the random walk model at the 5% level or above for the 730-day

[6]The estimation results of Nelson-Siegel model are available upon request.

implied volatility.

Among all the models, model (6), which runs VAR(1) on the volatility changes, has the greatest R^2_{OS} and performs the best. This implies that the historical information of other maturities is helpful when we forecast the implied volatility of one particular maturity. Model (7) and model (8), which use ECM models with common trends, also perform well. It is interesting to observe that most of the R^2_{OS} using Nelson-Siegel factors (model 1 and model 2) have negative values, suggesting that they are not as good as the benchmark model out-of-sample. This shows the difference of in-sample fitting and out-of-sample forecasts. In general, short-maturity implied volatilities tend to be more predictable than long-maturity implied volatilities. For example, when the VAR(1) on volatility change is used for the forecast, the R^2_{OS} of the 30-day implied volatility is 5.20%, while it is only 3.11% for the 730-day implied volatility.

It is also interesting to observe that overall the use of whole implied volatility curve information helps improve the forecasting performance, different models generate different results. For example, most of the R^2_{OS} of model (5), which runs VAR(1) on the volatility levels, are not significant. This is quite different from the results of model (6) that are highly significant. The results using the principal components of implied volatility curve (model (9) and model (10)) are not significant either. This finding is consistent with Kelly and Pruitt (2013, 2014) who find that the principal component will unfortunately contain the common error component that is not relevant to the forecasting, and has poor forecasting performance. Model (5), (9) and (10) use the level information, while model (6) to (8) use the information of volatility change and thus remove the trend. This performance difference implies both the information set and the way of modeling information set are important when when we run the out-of-sample forecast on the option market.

Turning now to the performance of the five-day forecast (the middle panel of Table 3), it is clear that most models perform worse than they do for the one-day forecast: they generally become less significant. Only model (15) has a positive R^2_{OS} that is significant at the 5% level for all maturities. In contrast, there are eight models meeting this criterion for the one-day forecast. In addition, the performance becomes much worse for long-maturity implied volatilities. There is only one model that is significant at the 5% level for the 730-day implied volatility. Model (6), model (7) and model (8) continue to perform quite well for the five-day forecast. The combination forecast (model 11 to model 16) seems to give stable and significant results. This means that the implied volatility is still predictable five days ahead when we use the daily data.

The bottom panel of Table 3 reports the results of the R^2_{OS} of the 20-day forecast. Most of the R^2_{OS} are insignificant. The forecasting abilities almost disappear when we forecast 20 days ahead, and none is able to generate a positive R^2_{OS} consistently across maturities that are significant at the 5% level. Models (6), (7) and (8) that perform well in the one-day and five-day forecast fail to beat the benchmark model in the 20-day forecast.

[Insert Table 3 here]

In order to visually observe the performance of models over time, we also calculate their monthly aggregate out-of-sample forecast errors and compare them with those of the random walk model. Figure 2 plots the difference of monthly aggregate out-of-sample forecast errors between model (6) (VAR (1) model on volatility change, the best-performing model reported in Table 3), and the random walk model. A negative value means the VAR (1) model on volatility change performs better in that month. We standardize the series to make the pattern clear. Figure 2 shows that for the one-day forecast of all maturity implied volatilities and the five-day forecast of short-maturity implied volatilities, most of the differences are negative, suggesting that model (6) consistently outperforms the random walk model during the sample period.

[Insert Figure 2 here]

B. Economic significance

The statistical significance results in Table 3 suggest that these models can forecast the implied volatilities, especially short-maturity implied volatilities, rather well up to five days. To explore the economic significance of this predictability, we further develop the option trading strategies as described in Section 2.2. We delta-hedge our option portfolio and include a transaction cost of $0.125 for each traded contract following Goncalves and Guidolin (2006). The stationary bootstrap with an average block size of 10 is applied to test the significance level of the Sharpe ratio and Leland's Alpha following Politis and Romano (1994) and Neumann and Skiadopoulos (2013). Since the benchmark model has a zero Sharpe ratio by design, any model with economically significant predictability would return a positive Sharpe ratio and Leland's Alpha.

Table 4 reports the results of the economic significance analysis. Model (6), (7) and (8) continue to perform well. The combination forecasts also provide better economic performance than the benchmark model. This suggests that they are both statistically and economically significant. The economic significance of the one-day forecast is much stronger than that of the five-day forecast. For model (6), which works the best, the Sharpe ratio and the Leland's Alpha for the one-day forecast is 1.75 and 34.92%, respectively. They decline to 0.38 and 11.52%, respectively, for the five-day forecast. None of these models is economically significant for the 20-day forecast. This is consistent with our finding that the historical implied volatility curve information is important to predict the implied volatilities up to one week. In summary, we find that the predictability of implied volatilities using the historical implied volatility curve information is both statistically and economically significant up to one week, and their predictability power lessens beyond a week.

[Insert Table 4 here]

Figure 3 plots the standardized aggregate monthly returns of the portfolios that are based on the forecast of implied volatility curve one day ahead. For those models that have a positive Sharpe ratio and Leland's Alpha (model 6, 7, 8, 11, 13, 14, 15 and 16), their returns are relatively stable during the normal time. The returns during the financial crisis become much more volatile. Most of them have a large downward spike in the crisis period, which suggests that these trading strategies could be subject to downside risk. The exception is model (6), which runs VAR(1) on volatility changes. It has a sudden return increase during the financial crisis period. Using model (6) provides a better hedge against the downside risk compared with other models. Figure 4 plots the standardized aggregate monthly returns of the portfolios that are based on the forecast of implied volatility curve five days ahead and the findings are similar.

[Insert Figure 3 here]

[Insert Figure 4 here]

IV. Robustness checks

A. Out-of-sample forecast during the recent financial crisis

Our data cover the recent financial crisis period. One interesting question is whether the predictability changes over that period. We examine the performance of out-of-sample forecasts between December 2007 and June 2009, the recession periods indicated by the National Bureau of Economic Research (NBER). Table 5 reports the results of statistical and economic significance. Panel A reports the results of R^2_{OS}, while Panel B reports the results of economic significance. Consistent with the results for the whole sample, the out-of-sample one-day forecast during the financial crisis is statistically significant for several models, although it becomes weaker. For example, when the daily data are used, only four models (model 6, 7, 8 and 16) can beat the benchmark model in forecasting the 30-day implied volatility one day ahead, which is less significant than the results using the whole sample. None of the five-day forecasts is significant. This finding is different from other research that finds stronger predictability during the recession period. One possible reason is that during the crisis period, the investors are more sensitive to the information on the financial markets. As a result, it takes less time for the option market to adjust for the information.

Panel B of Table 5 reports the results of economic significance. Different from the results of statistical significance, the economic significance of predictability, on the contrary, becomes stronger. For example, the option trading strategy using one-day forecasts based on model (6) could generate a Sharpe ratio as high as 2.80 and a 95.86% Leland's Alpha. There are seven models that have Sharpe ratios greater than one even after transaction costs are considered. These results are much stronger than those in Table 4. This implies that the historical information is more economically important during the crisis period. Our finding is consistent with Loh and Stulz (2014), who find that analysts tend to make poor forecasts during crisis, but the forecasts become more influential once they are adjusted.

[Insert Table 5 here]

B. Out-of-sample forecast using weekly data

To support our findings from another perspective, we re-arrange data to weekly frequency by using observations on Wednesday because the number of holidays on Wednesdays is the least among the five weekdays (Li and Zhang, 2010). Table 6 reports the results of out-of-sample forecasts. For simplicity, we only report the R^2_{OS} of all maturities.[7] The results continue to show the predictability of implied volatilities one week ahead. There are 10 models that have significant R^2_{OS}, with the highest being 5.01% by model (6). Four models

[7]The R^2_{OS} of other maturities are available upon request.

(model 6, 7, 8 and 16) generate a significant Sharpe ratio and Leland's Alpha. However, they can barely outperform the random walk model beyond a one-week forecasting horizon. The highest R^2_{OS} are only 0.28% and 0.72% for two-week forecasts and four-week forecasts respectively. None of the economic profit is significant beyond one week.

[Insert Table 6 here]

C. Out-of-sample forecast of other option series

Another question of interest is whether the findings using the ATM call options could be extended to other option series. In order to answer this question, we try three different option series, including the ATM put options, the call options with delta equal to 0.60 and the call options with delta equal to 0.40. Table 7 reports the results of these option series.

The left column of Table 7 reports the out-of-sample forecast of ATM put options. The up panel reports the results of R^2_{OS} and the bottom panel reports the results of Leland's Alpha. For simplicity, we only report the best results of the 16 models. The numbers in the bracket represent the number of models that outperform the benchmark model at least at the 10% significance level. The historical implied volatility curve information is important up to 20 days during the whole sample period. Short-maturity implied volatilities tend to be more predictable. For example, for the one-day forecast, the highest R^2_{OS} of a 30-day ATM put option is 2.03% with 10 models outperforming the benchmark model, while it is only 0.87% with only one model outperforming the benchmark model for the 730-day ATM put option. The results in the financial crisis period show weaker predictability for the ATM put options. Now they are only predictable one day ahead.[8] The results of Leland's Alpha also show that the predictability of the implied volatilities of ATM put options is also of economic significance. The trading strategy using the five-day forecast could generate a

[8]We also tried the weekly data and found that the predictability disappears when the weekly data are used.

Leland's Alpha of 13.04% during the whole period, and 34.55% during the financial crisis period. Although the predictability becomes weaker during the financial crisis period, its economic value is much higher.

The middle column and the right column of Table 7 report the results of call options with delta equal to 0.60 and 0.40, respectively. The results are quite similar to the ATM put options. The implied volatilities could be predicted up to 20 days in the whole period. The predictability is also economically significant. The predictability becomes weaker during the financial crisis, but the economic value is much higher. The importance of historical information lasts longer for the ATM put options and for the call options with other moneyness than for the ATM call options. One possible reason is that the trading of other options is not as liquid as the ATM call options and therefore it takes a longer time for them to reflect the historical information.

[Insert Table 7 here]

V. Price Discovery Analysis

The above empirical analysis shows that the whole implied volatility curve provides useful information on the forecasting of term structure of implied volatilities. It is therefore interesting to investigate which maturity's implied volatility contain more information than the others. We use the Hasbrouck (1995) and Gonzalo-Granger (1995) price discovery measures to assess the information share of implied volatilities. In order to control for the leverage, we run the price discovery analysis among implied volatilities with the same moneyness.

Table 8 reports the results of price discovery analysis. The VECM models are used to calculate the price discovery measures. The left panel reports the results of the cointegration relationship. There exists a significant cointegration relationship for all pairs. The middle panel reports the results of the Hasbrouck measure, while the right panel reports the results of the Gonzalo-Granger measure. In the middle and right panel, the number in maturity row i and maturity column j is the price discovery measure of the i-day implied volatility when the VECM model is run between the i-day implied volatility and the j-day implied volatility. For example, the number with maturity row 30 and column 91 in the middle panel is 0.44. It is the Hasbrouck measure of the 30-day implied volatility when the VECM model is run using the 30-day and 91-day implied volatilities. In order to measure the aggregate level of information share of each maturity implied volatility, we calculate the mean price discovery measure of one maturity option using the mean of its four individual price discovery measures.

The results in Table 8 strongly indicate that long-maturity implied volatilities have larger price discovery measures compared with short-maturity implied volatilities. The price discovery measures of the 365-day implied volatility are the largest among all maturities in most cases. They are also much larger than those of the 30-day implied volatility. For example, for the ATM call options, the mean Hasbrouck measure and the Gonzalo-Granger measure of the 365-day implied volatility are 0.56 and 0.93, respectively, while those of the 30-day implied volatility are only 0.41 and 0.06, respectively. The implied volatility of the 365-day option contains more useful information for the price discovery of the term structure of implied volatilities. The results for the ATM put option and call options with other moneyness are quite similar. Figure 5 plots the mean price discovery measure of the implied volatilities. The findings are also consistent with Table 8.

[Insert Table 8 here]

[Insert Figure 5 here]

Next we examine whether the larger price discovery measures of long-maturity implied volatilities are due to the larger trading volume of long-maturity options. Table 9 reports the trading summary of options with different maturities. We report both the trading volume and the open interest. The option data with negative bid-ask spread, negative trading volume and open interest or negative implied volatility are excluded. The trading volume and open interest of ATM (call and put) options, call options with delta equal to 0.60 and call options with delta equal to 0.40 are calculated from the options with moneyness between 45% and 55%, between 55% and 65% and between 35% and 45%, respectively.

The trading of the option market is dominated by short-maturity options. For example, for the ATM call options, the options with maturity less than three months contribute about 78.35% to the total trading volume and about 53.96% to the total open interest. On the other hand, the options with maturity longer than one year only account for 3.16% of the total trading volume and 9.88% of the total open interest. The trading of long-maturity options is much less than that of short-maturity options.

The results in Table 8 and Table 9 together suggest that although long-maturity options are not much traded, their prices contain much more information. One possible reason is that much trading of short-maturity options is noise trading and tends to be more affected by behaviour bias. On the other hand, the trading of long-maturity options is more information driven, which helps price discovery in a more efficient way. Our findings are not consistent with Chakravarty, Gulen and Mayhew (2004) about the positive relationship between price discovery and liquidity on the option market once the leverage is controlled, but consistent with Barclay and Hendershott (2003) and Zhu (2014) who find that high level of price discovery could happen on a less liquid market.

[Insert Table 9 here]

VI. Conclusion

In this paper, we test the out-of-sample predictability of the term structure of S&P500 index implied volatilities. In particular, we evaluate 16 different models that are based on historical implied volatility information. We investigate both their statistical and economic significance. We run several robustness tests using the sub-period data and the data of different option series. In order to examine how long this predictability lasts, we also compare the results at different horizons. Finally, we assess the information share of implied volatilities. We obtain several interesting results.

Using out-of-sample R^2_{OS} as the statistical measure, we find that several models that use the historical implied volatility curve information could predict the term structure of implied volatilities significantly out-of-sample. When the daily data are used, these models could forecast the implied volatilities up to five days ahead. Short-maturity implied volatilities tend to be more predictable than long-maturity implied volatilities.

Using the Sharpe ratio and Leland's Alpha as the economic significance measures, we find that the predictability is of economic significance. The models that use the information of implied volatility curve generate a positive Sharpe ratio and Leland's Alpha, even if the transaction cost is accounted for. The portfolio returns that are based on these models are relatively stable during normal times but much more volatile during the financial crisis period. The VAR(1) model on volatility change performs quite well in hedging against the downside risk of the crisis. Using the data during the financial crisis, we find that overall, the predictability becomes weaker during the financial crisis, but the economic value is much higher. The robustness tests that use other option series also support our main findings.

Using the Hasbrouck and Gonzalo-Granger price discovery measures, we find that although long-maturity options are much less traded than short-maturity options, their implied volatilities contain more useful information for the price discovery of the term structure of implied volatilities.

These findings are relevant to the question of how long it takes for option prices to reflect the historical price information. Our results show the importance of historical information up to one week for the ATM call options. For the ATM put options and the call options with other moneyness, the historical information could be helpful up to 20 days. This question is important for the understanding of the Efficient Market Hypothesis (EMH) in practice and for the policy implementation to reduce the time interval.

Our analysis of price discovery also contributes to understanding the relationship between liquidity and price discovery. The findings about the information share of long-maturity options also have implications about the future studies of option pricing models. Our empirical results are consistent with the emerging component volatility models. In other words, both short-term and long-term volatilities should be considered in the option pricing models to fully utilize the information from the term structure of implied volatilities.

References

Bakshi, G.; C. Cao; and Z. Chen. "Empirical Performance of Alternative Option Pricing Models." *Journal of Finance*, 52 (1997), 2003–2049.

Barclay, M. J., and T. Hendershott. "Price Discovery and Trading After Hours". *Review of Financial Studies* 16 (2003), 1041–1073.

Bodurtha, J. N., and G. R. Courtadon. "Efficiency Tests of the Foreign Currency Options Market." *Journal of Finance*, 41 (1986), 151–162.

Chalamandaris, G., and A. E. Tsekrekos. "Predictable Dynamics in Implied Volatility Surfaces from OTC Currency Options." *Journal of Banking and Finance*, 34 (2010), 1175–1188.

Chalamandaris, G., and A. E. Tsekrekos. "How Important is the Term Structure in Implied Volatility Surface Modeling? Evidence from Foreign Exchange Options." *Journal of International Money and Finance*, 30 (2011), 623–640.

Chakravarty, S.; H. Gulen; and S. Mayhew. "Informed Trading in Stock and Option Markets." *Journal of Finance*, 59 (2004), 1235–1257.

Chiras, D. P., and S. Manaster. "The Information Content of Option Prices and a Test of Market Efficiency." *Journal of Financial Economics*, 6 (1978), 213–234.

Christoffersen, P.; S. Heston; and K. Jacobs. "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well?" *Management Science*, 55 (2009), 1914–1932.

Christoffersen, P.; K. Jacobs; C. Ornthanalai; and Y. Wang. "Option Valuation with Long-run and Short-run Volatility Components." *Journal of Financial Economics*, 90 (2008), 272–297.

Derman, E.; I. Kani; and J. Z. Zou. "The Local Volatility Surface: Unlocking the Information in Index Option Prices." *Financial Analysts Journal*, 52 (1996), 25–36.

Diebold, F. X., and C. Li. "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics*, 130 (2006), 337–364.

Galai, D. "Tests of Market Efficiency of the Chicago Board Options Exchange." *Journal of Business*, 50 (1977), 167–197.

Goncalves, S., and M. Guidolin. "Predictable Dynamics in the S&P 500 Index Options Implied Volatility Surface." *Journal of Business*, 79 (2006), 1591–1635.

Gonzalo, J., and C. Granger. "Estimation of Common Long Memory Components in Cointegrated Systems." *Journal of Business and Economics Statistics*, 13 (1995), 27–35.

Hansen, B. E. "Notes and Comments: Least Squares Model Averaging." *Econometrica*, 75 (2007), 1175–1189.

Hansen, B. E. "Least-squares Forecast Averaging." *Journal of Econometrics*, 146 (2008), 342–350.

Harvey, C. R., and R.E. Whaley. "Market Volatility Prediction and the Efficiency of the S&P 100 Index Option Market." *Journal of Financial Economics*, 31 (1992), 43–73.

Hasbrouck, J. "One Security, Many Markets: Determining the Contribution to Price Discovery." *Journal of Finance*, 50 (1995), 1175-1199.

Hotchkiss, E. S., and T. Ronen. "The Informational Efficiency of the Corporate Bond Markets: An Intraday Analysis." *Review of Financial Studies*, 15 (2002), 1325–1354.

Hodrick, R. J. "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement." *Review of Financial Studies*, 5 (1992), 357–386.

Hogan, S.; R. Jarrow; M. Teo; and M. Warachka. "Testing Market Efficiency Using Statistical Arbitrage with Applications to Momentum and Value Strategies." *Journal of Financial Economics*, 73 (2004), 525–565.

Kelly, B., and S. Pruitt. "Market Expectations in the Cross-section of Present Values." *Journal of Finance*, 68 (2013), 1721–1757.

Kelly, B., and S. Pruitt. "The Three-pass Regression Filter: A New Approach to Forecasting Using Many Predictors." *Journal of Econometrics*, forthcoming.

Klemkosky, R. C., and B. G. Resnick. "Put–call Parity and Market Efficiency." *Journal of Finance*, 34 (1979), 1141–1155.

Konstantinidi, E.; G. Skiadopoulos; and E. Tzagkaraki. "Can the Evolution of Implied Volatility be Forecasted? Evidence from European and US Implied Volatility Indices." *Journal of Banking and Finance*, 32 (2008), 2401–2411.

Kwan, S. H. "Firm-specific Information and the Correlation between Individual Stocks and Bonds." *Journal of Financial Economics*, 40 (1996), 63–80.

Leland, H. E. "Beyond Mean-variance: Performance Measurement in a Nonsymmetrical World." *Financial Analysts Journal*, 55 (1999), 27–36.

Li, G., and C. Zhang. "On the Number of State Variables in Options Pricing." *Management Science*, 55 (2010), 2058–2075.

Loh, R. K., and R. M. Stulz. "Is Sell-side Research More Valuable in Bad Times?" Working paper, NBER (2014).

Nelson, C. R., and A. F. Siegel. "Parsimonious Modelling of Yield Curves." *Journal of Business*, 60 (1987), 473–489.

Neumann, M., and G. Skiadopoulos. "Predictable Dynamics in Higher Order Risk-neutral Moments: Evidence from the S&P 500 Options." *Journal of Financial and Quantitative Analysis*, 48 (2013), 947–977.

Phillips, S. M., and C. W. Smith. "Trading Costs for Listed Options: The Implications for Market Efficiency." *Journal of Financial Economics*, 8 (1980), 179–201.

Politis, D. N., and J. P. Romano. "The Stationary Bootstrap." *Journal of the American Statistical Association*, 89 (1994), 1303–1313.

Poon, S. H., and P. Pope. "Trading Volatility Spreads: A Test of Index Option Market Efficiency." *European Financial Management*, 6 (2000), 235–260.

Rapach, D. E.; J. K. Strauss; and G. F. Zhou. "Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23 (2010), 821–862.

Zhu, H. "Do Dark Pools Harm Price Discovery?" *Review of Financial Studies*, 27 (2014), 747–789.

Table 1. Prediction models
This table lists the 16 different prediction models to be tested in this paper. The last row also explains the benchmark model (random walk), which assumes that the historical information is not useful in the predictions and uses the current value as the best forecast.

Model	Model description
1	Nelson-Siegel factors as univariate AR(1) processes
2	Nelson-Siegel factors as multivariate VAR(1) processes
3	Slope regression
4	AR(1) on volatility levels
5	VAR(1) on volatility levels
6	VAR(1) on volatility changes
7	ECM(1) with one common trend
8	ECM(1) with two common trends
9	AR(1) regression on three principal components
10	VAR(1) on empirical level, slope and curvature
11	mean combination forecast
12	median combination forecast
13	trimmed mean combination forecast
14	DMSPE combination forecast with $\theta = 1$
15	DMSPE combination forecast with $\theta = 0.9$
16	MMA combination forecast
Benchmark	Random walk

Table 2. Summary statistics

This table reports the summary statistics of implied volatilities used in the empirical analysis. The sample period is from 1996 to 2011. The maturities include 30 days, 60 days, 91 days, 122 days, 152 days, 182 days, 273 days, 365 days, 547 days and 730 days. This table reports the mean, standard deviation, minimum, maximum, autocorrelation with lag 10, 30, 60 and 180 days of these implied volatilities.

Maturity (days)	Mean (%)	Std. Dev. (%)	Min. (%)	Max. (%)	ρ (10)	ρ (30)	ρ (60)	ρ (180)
30	20.07	7.81	8.32	74.98	0.89	0.74	0.58	0.28
60	20.20	7.13	9.07	67.27	0.92	0.80	0.64	0.32
91	20.26	6.72	9.68	60.66	0.93	0.82	0.67	0.35
122	20.29	6.34	10.22	57.50	0.94	0.84	0.70	0.38
152	20.32	6.05	10.44	53.99	0.95	0.85	0.72	0.41
182	20.33	5.84	10.59	50.44	0.95	0.86	0.74	0.43
273	20.34	5.48	10.96	46.49	0.96	0.88	0.78	0.47
365	20.36	5.28	11.25	44.49	0.96	0.89	0.79	0.49
547	20.40	4.93	11.61	40.19	0.97	0.91	0.82	0.52
730	20.45	4.78	11.74	38.39	0.97	0.91	0.82	0.53

Table 3. R^2_{OS} of implied volatility forecast

This table reports the R^2_{OS} of implied volatility forecast by the 16 prediction models. The out-of-sample forecast starts in 2002. When calculating the R^2_{OS}, the random walk model is used as the benchmark. A positive R^2_{OS} indicates the prediction model outperforms the benchmark model. The statistical significance for the R^2_{OS} statistics is based on the p-value of the MSPE-adjusted statistics. Hodrick (1992) standard error is used for p-value calculation to account for the impact of overlapping residuals. ^{a}, ^{b}, ^{c} denote significance at the 1%, 5% and 10% level, respectively.

	Model															
Maturity	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	One day ahead (%)															
30	-0.20	0.74^b	-0.03	-0.21	0.44^b	5.20^a	5.10^a	5.34^a	1.69^b	0.78^b	4.18^a	3.07^a	4.03^a	4.24^a	4.18^a	5.10^a
91	-8.91	-3.05	0.81^c	0.02^c	0.13^c	4.97^a	3.73^a	4.16^a	-2.66	-3.97	2.22^a	1.82^a	2.14^a	2.41^a	2.37^a	4.33^a
152	-5.90	-1.70	0.62^c	0.02^c	-0.25	4.55^a	3.82^a	4.24^a	-2.67	-4.42	1.94^a	1.54^a	1.89^a	2.03^a	1.96^a	3.77^a
365	-17.17	-5.51	0.08	-0.03	0.23^c	4.56^a	4.02^a	4.21^a	-2.90	0.16^c	2.43^a	2.14^a	2.56^a	2.58^a	2.56^a	3.95^a
730	-16.08	-12.30	-0.26	-0.18	-0.54	3.11^a	2.37^a	2.50^a	-14.47	-50.35	-0.78	0.87^a	0.07^b	0.14^a	0.19^b	2.51^a
All	-4.40	-1.21	0.24^b	-0.12	0.24^a	4.96^a	4.47^a	4.78^a	-0.63	-2.60	3.19^a	2.47^a	3.12^a	3.31^a	3.26^a	4.60^a
	Five days ahead (%)															
30	1.49^b	1.41^c	-0.12	0.14^b	0.65^b	8.14^a	6.80^a	7.36^a	3.33^b	0.40^c	5.51^a	4.19^b	5.13^a	5.64^a	6.44^a	8.42^a
91	-6.66	-3.94	0.37	0.083^b	-1.81	4.81^b	3.97^a	4.80^b	-1.75	-2.86	2.01^b	0.75^c	1.75^b	2.17^b	3.01^b	4.82^a
152	-3.25	-0.84	-0.61	-0.04	-1.97	4.35^b	3.52^b	4.33^b	0.83^b	-1.64	2.32^b	1.09^c	2.01^b	2.40^b	3.08^a	4.64^a
365	-9.79	-2.29	-1.09	-0.19	-2.40	3.22^b	1.86^c	2.70^c	-0.26	-1.78	1.64^c	0.57	1.29	1.68^c	2.62^b	3.04^a
730	-7.73	-6.64	-1.25	-0.50	-3.13	1.23^c	-0.48	0.46	-7.05	-16.78	-0.41	-0.82	-0.45	-0.24	1.21^b	0.87
All	-2.19	-0.72	-0.19	0.05^a	-0.64	6.23^a	5.06^a	5.75^a	1.15^a	-1.48	3.75^a	2.51^a	3.43^a	3.88^a	4.71^a	6.39^a
	20 days ahead (%)															
30	1.86^a	-0.01	-0.58	2.10^a	-2.13	-4.20	-1.00	-0.67	3.65^a	-1.30	3.27^b	1.07^b	2.65^b	3.44^b	6.85^b	5.57
91	-5.18	-4.22	-3.30	0.11^a	-5.00	-4.34	-4.16	-4.23	0.26^a	-3.68	-0.33	-2.50	-0.97	-0.11	2.87^a	0.35
152	-5.79	-3.89	-5.33	-0.78	-6.16	-6.13	-5.19	-5.05	0.82^a	-4.72	-1.52	-3.20	-2.18	-1.32	1.28^b	-0.16
365	-10.07	-4.36	-5.25	-1.15	-6.74	-7.32	-6.87	-7.07	0.78^a	-5.48	-2.15	-3.89	-2.74	-2.01	1.06	-0.23
730	-7.70	-4.66	-3.53	-1.48	-5.11	-8.22	-7.89	-7.46	-1.07	-6.82	-1.55	-2.82	-1.90	-1.42	2.61^a	-0.14
All	-2.41	-2.18	-2.47	0.75^a	-3.94	-4.98	-3.21	-3.04	1.94^a	-3.00	1.01^a	-1.02	0.39^a	1.19^a	4.38^a	2.71

Table 4. Economic significance of implied volatility forecast

This table reports the Sharpe ratio and Leland's (1999) Alpha of each model. On date t, we long (short) an option if the forecasted volatility for that maturity at date $t+h$ is larger (smaller) than the current volatility. We consider options with maturity at 30, 91, 152, 365 and 730 days; thus on each date our portfolio includes five option contracts with trading status either long, short or no trading. We hold the portfolio the same time as our forecasting horizon and repeat the trading in the out-of-sample period. We delta-hedge our option portfolio and include a transaction cost of \$0.125 for each traded contract. The stationary bootstrap with an average block size of 10 is applied to test the significance level. ^{a}, ^{b}, ^{c} denote significance at the 1%, 5% and 10%, level respectively.

	Model															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	One day ahead															
Sharpe ratio	-0.49	-0.30	-0.25	-0.35	-0.17	1.75^a	0.82^a	0.81^a	-0.73	-0.55	0.31^a	-0.34	0.44^a	0.32^a	0.23^a	1.35^a
Leland's Alpha (%)	–10.99	-4.95	-3.39	-6.71	-1.09	34.92^a	24.69^a	24.48^a	-19.5	-13.13	12.38^a	-1.32	15.60^a	12.99^a	10.50^a	35.17^a
	Five days ahead															
Sharpe ratio	-0.18	-0.11	-0.39	-0.28	-0.22	0.38^a	0.33^a	0.33^a	-0.06	-0.29	0.10^a	0.06^a	0.07^a	0.10^a	0.13^a	0.36^a
Leland's Alpha (%)	-1.82	-0.12	-7.24	-4.25	-2.98	11.52^a	10.46^a	10.37^a	1.12	-4.52	5.07^a	4.52^a	4.31^a	5.24^a	5.79^a	11.53^a
	20 days ahead															
Sharpe ratio	-0.20	-0.04	-0.51	-0.20	-0.11	-0.61	-0.26	-0.22	-0.07	-0.03	-0.13	-0.12	-0.12	-0.11	-0.05	-0.01
Leland's Alpha (%)	-2.41	1.39	-9.30	-2.30	-0.14	-11.63	-3.72	-2.65	0.60	1.58	-0.66	-0.37	-0.57	-0.26	1.09	2.27

Table 5. Predictability of implied volatility during the financial crisis

This table reports the predictability of implied volatility during the financial crisis period. Panel A reports the results of R^2_{OS}, while Panel B reports the results of economic significance. When calculating the R^2_{OS}, the random walk model is used as the benchmark. 16 models are used in the evaluation. A positive R^2_{OS} indicates the prediction model outperforms the benchmark model. The statistical significance for the R^2_{OS} statistics is based on the p-value of the MSPE-adjusted statistics. Hodrick (1992) standard error is used for p-value calculation to account for the impact of overlapping residuals. Panel B reports the Sharpe ratio and Leland's (1999) Alpha of each model during the recent financial crisis. On date t, we long (short) an option of the maturity if the forecasted volatility for that maturity at date $t + h$ is larger (smaller) than the current volatility. We consider options with maturity at 30, 91, 152, 365 and 730 days; thus on each date our portfolio includes five option contracts with trading status either long, short or no trading. We hold the portfolio the same time as our forecasting horizon and repeat the trading in the out-of-sample period. We delta-hedge our option portfolio and include a transaction cost of \$0.125 for each traded contract. The stationary bootstrap with an average block size of 10 is applied to test the significance level. ^{a}, ^{b}, ^{c} denote significance at the 1%, 5% and 10% level, respectively.

Panel A. R^2_{OS}

	Model															
Maturity	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	One day ahead (%)															
30	-0.98	-0.57	-0.03	-0.04	-1.48	5.37^b	3.65^c	4.58^b	0.90	-0.45	3.41	2.27	3.31	3.48	3.34	4.23^a
91	-2.63	0.88	-0.39	-0.02	0.16	6.39^a	4.51^b	4.65^b	0.66	-0.72	3.34^c	2.60^c	3.08^c	3.44^c	3.25^c	5.42^a
152	-8.48	-5.01	-0.11	-0.14	-1.36	6.39^a	4.41^b	5.59^a	-4.95	-7.41	1.19	0.83	1.16	1.36	1.27	5.21^a
365	-23.06	-10.68	-0.57	-0.17	-1.14	7.37^a	5.11^b	5.46^b	-1.58	-1.51	2.14	1.94	2.40^c	2.39^c	2.25	5.42^a
730	-10.86	-7.92	-1.14	-0.31	-1.89	5.43^b	2.89^b	2.98^b	-12.68	-52.04	0.87	2.12^c	1.31	1.66	1.32	3.81^a
All	-3.75	-1.52	-0.18	-0.06	-1.08	5.85^a	4.01^a	4.73^a	-0.36	-2.74	2.99^b	2.15^c	2.89^b	3.10^b	2.95^b	4.68^a
	Five days ahead (%)															
30	0.31	-1.10	-0.11	-0.13	-3.55	7.40	1.75	4.91	2.03	-3.02	3.45	1.99	2.99	3.56	4.35	6.79
91	-5.77	-5.73	-2.82	-0.41	-4.40	3.55	1.26	3.13	-1.90	-5.88	0.25	-1.42	-0.21	0.36	0.94	3.33
152	-5.86	-4.85	-3.67	-0.91	-5.74	3.31	0.40	3.31	0.21	-6.71	0.26	-1.38	-0.25	0.33	1.10	3.38
365	-18.60	-9.34	-3.89	-1.12	-7.30	2.47	-2.31	-0.28	0.96	-7.45	-0.92	-2.44	-1.54	-0.87	0.15	1.70
730	-6.38	-9.96	-3.83	-1.57	-7.45	-0.99	-4.26	-2.76	-7.88	-24.24	-1.39	-2.74	-1.88	-1.26	1.03	-0.89
All	-3.13	-3.35	-1.50	-0.39	-4.35	5.48^c	1.07	3.79	0.56	-5.00	1.92	0.39	1.45	2.03	2.82	5.06^c
	20 days ahead (%)															
30	0.48	-3.23	-0.60	1.99	-6.53	-11.87	-5.86	-5.46	0.82	-5.62	-0.46	-2.55	-1.34	-0.34	2.09	2.91
91	-5.39	-6.79	-9.84	-0.06	-7.49	-8.30	-7.85	-8.13	-0.99	-7.60	-3.78	-5.83	-4.84	-3.72	-2.21-5.57	
152	-8.45	-7.86	-12.82	-2.29	-10.56	-10.93	-10.88	-10.55	0.31	-10.18	-5.90	-7.31	-7.05	-5.81	-4.21	-6.08
365	-17.36	-10.34	-10.84	-3.32	-12.28	-9.38	-15.59	-15.32	2.68	-11.87	-7.08	-8.89	-7.92	-6.96	-5.00	-3.11
730	-9.11	-5.45	-6.03	-4.35	-7.32	-10.37	-15.92	-13.72	6.05^b	-6.29	-3.12	-5.72	-4.20	-3.03	0.44	2.24
All	-3.82	-5.32	-5.48	0.27	-7.77	-10.66	-8.13	-7.84	0.65	-7.21	-2.60	-4.58	-3.56	-2.50	-0.40	-0.84

Panel B. Economic significance

	Model															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
One day ahead																
Sharpe ratio	-0.53	-0.22	-0.13	-0.61	-0.32	2.80[a]	1.04[a]	0.84[a]	-1.10	-0.91	1.36[a]	-0.14	1.48[a]	1.24[a]	1.09[a]	2.64[a]
Leland's Alpha (%)	-23.38	-6.98	-2.20	-32.90	-13.08	95.86[a]	47.91[a]	41.02[a]	-70.73	-53.13	58.79[a]	-2.84	61.45[a]	54.55[a]	50.06[a]	91.65[a]
Five days ahead																
Sharpe ratio	-0.33	-0.11	-0.44	-0.71	-0.39	0.37[a]	0.26[a]	0.58[a]	-0.36	-0.53	0.25[a]	0.07[a]	0.17[a]	0.28[a]	0.26[a]	0.52[a]
Leland's Alpha (%)	-8.47	0.56	-13.86	-27.31	-11.32	21.39[a]	17.48[a]	27.45[a]	-10.61	-16.35	14.66[a]	8.87[a]	12.72[a]	15.83[a]	15.15[a]	26.83[a]
20 days ahead																
Sharpe ratio	-0.41	-0.35	-0.65	-0.79	-0.29	-0.82	-0.29	-0.22	-0.70	-0.44	-0.32	-0.32	-0.32	-0.31	-0.29	-0.56
Leland's Alpha (%)	-10.07	-7.57	-19.49	-27.43	-4.71	-27.23	-4.73	-2.15	-23.87	-11.41	-6.25	-6.38	-6.19	-6.17	-5.10	-15.96

Table 6. Out-of-sample forecast of implied volatilities of ATM call options: Weekly data.

This table reports the out-of-sample forecast of implied volatilities of ATM call options by the 16 models using weekly data. The up, middle and bottom panels report the results of forecast one week ahead, two weeks ahead and four weeks ahead, respectively. For simplicity, we only report the R^2_{OS} of all maturities. The statistical significance for the R^2_{OS} statistics is based on the p-value of the MSPE-adjusted statistics. Hodrick (1992) standard error is used for p-value calculation to account for the impact of overlapping residuals. When calculating the Sharpe ratio and Leland's Alpha, we delta-hedge our option portfolio and include a transaction cost of \$0.125 for each traded contract. The stationary bootstrap with an average block size of 10 is applied to test the significance level. ^{a}, ^{b}, ^{c} denote significance at the 1%, 5% and 10% level, respectively.

All	Model															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	One week ahead															
R^2_{OS} (%)	-3.22	-1.38	-0.43	-0.55	-1.34	5.01^a	2.77^a	3.82^a	0.34^b	-2.37	2.50^a	1.48^b	2.03^a	2.54^a	2.49^a	2.30^a
Sharpe ratio	-0.26	-0.21	-0.64	-0.21	-0.45	0.37^a	0.16^a	0.25^a	-0.12	-0.49	-0.19	-0.11	-0.15	-0.11	-0.12	0.09^a
Leland's Alpha (%)	-3.51	-2.38	-13.02	-2.38	-8.14	10.32^a	5.83^a	7.81^a	-0.34	-9.03	-1.89	-0.15	-1.00	-0.20	-0.46	4.47^a
	Two weeks ahead															
R^2_{OS} (%)	-3.35	-3.07	-1.49	-0.98	-4.28	-3.79	-3.38	-2.72	0.28^a	-3.48	-0.14	-1.32	-0.53	-0.09	0.23	-2.19
Sharpe ratio	-0.35	-0.21	-0.63	-0.28	-0.39	-0.45	-0.51	-0.28	-0.27	-0.29	-0.36	-0.37	-0.33	-0.4	-0.39	-0.26
Leland's Alpha (%)	-5.63	-2.51	-11.34	-3.95	-6.35	-7.64	-9.18	-3.96	-3.69	-4.04	-5.85	-6.13	-5.27	-7.27	-6.88	-3.81
	Four weeks ahead															
R^2_{OS} (%)	-3.91	-3.71	-3.08	-0.56	-5.84	-6.72	-5.07	-5.24	0.72^a	-4.64	-0.46	-2.22	-0.98	-0.33	0.71^b	-0.70
Sharpe ratio	-0.25	-0.11	-0.63	-0.16	-0.23	-0.70	-0.50	-0.28	-0.08	-0.04	-0.18	-0.16	-0.19	-0.24	-0.22	-0.16
Leland's Alpha (%)	-3.32	-0.35	-10.92	-1.45	-2.75	-12.87	-8.35	-3.76	0.15	1.23	-1.86	-1.41	-2.29	-3.47	-3.08	-1.03

Table 7. Out-of-sample forecast of implied volatilities of other option series: A robustness check.

This table reports the out-of-sample forecast of implied volatilities of other option series by the 16 models. The left, middle and right columns report the results of ATM put options, call options with delta equal to 0.60 and call options with delta equal to 0.40, respectively. For simplicity, we only report the best results by the 16 prediction models. The number in the bracket denotes the number of models that outperform the benchmark model with at least 10% significance level. ^{a}, ^{b}, ^{c} denote significance at the 1%, 5% and 10% level, respectively.

	Maturity	ATM put options			Call options with delta = 0.60			Call options with delta = 0.40		
	(days)	One day ahead	Five days ahead	20 days ahead	One day ahead	Five days ahead	20 days ahead	One day ahead	Five days ahead	20 days ahead
			R^2_{OS}			R^2_{OS}			R^2_{OS}	
Whole period	30	$2.03^a(10)$	$6.92^a(11)$	$6.48^a(9)$	$6.20^a(14)$	$8.96^b(15)$	$6.89^b(5)$	$4.93^a(14)$	$8.66^a(15)$	$6.78^a(8)$
	91	$1.54^a(4)$	$4.20^a(5)$	$3.42^a(2)$	$5.04^a(9)$	$5.17^b(7)$	$3.72^b(3)$	$4.22^a(10)$	$5.40^a(8)$	$4.03^a(5)$
	152	$1.23^a(4)$	$3.27^a(9)$	$0.32^a(1)$	$5.14^a(11)$	$5.06^c(10)$	$1.35^a(1)$	$3.70^a(9)$	$5.66^a(11)$	$1.90^a(2)$
	365	$1.50^a(9)$	$2.09^a(7)$	$0.39^a(2)$	$5.03^a(9)$	$3.38^c(3)$	$1.30^a(1)$	$3.71^a(10)$	$3.46^a(6)$	$1.17^c(2)$
	730	$0.87^a(1)$	$1.14^a(2)$	-0.28(0)	$3.91^a(9)$	$1.09^c(1)$	$2.69^a(1)$	$2.03^a(5)$	$1.90^a(7)$	$3.10^a(1)$
	All	$1.35^a(9)$	$5.10^a(10)$	$3.77^a(5)$	$7.41^a(9)$	$6.87^a(11)$	$4.77^a(6)$	$4.19^a(10)$	$6.80^a(10)$	$4.74^a(7)$
Financial crisis period	30	0.80(0)	$3.30^b(2)$	$0.61^b(1)$	$6.01^b(4)$	7.47(0)	2.19(0)	$4.94^b(4)$	$7.11^c(1)$	$0.59^c(1)$
	91	$2.18^b(1)$	1.59(0)	0.17(0)	$6.01^a(3)$	3.27(0)	0.20(0)	$5.85^a(3)$	3.88(0)	-0.38(0)
	152	$1.78^c(1)$	0.39(0)	-2.32(0)	$6.40^a(4)$	3.75(0)	0.68(0)	$5.61^a(4)$	$4.20^c(1)$	1.09(0)
	365	$2.43^b(1)$	2.76(0)	$3.08^b(1)$	$8.11^a(5)$	2.36(0)	3.09(0)	$6.21^a(4)$	2.69(0)	3.01(0)
	730	$2.29^c(1)$	-0.26(0)	$0.31^c(1)$	$5.67^a(4)$	-0.11(0)	5.10(0)	$3.82^c(4)$	1.72(0)	$3.80^c(1)$
	All	$1.42^b(2)$	$2.12^a(2)$	-0.78(0)	$6.16^a(8)$	5.48(0)	1.13(0)	$5.27^a(8)$	$5.56^b(2)$	$0.22^c(1)$
			Leland's Alpha (%)			Leland's Alpha (%)			Leland's Alpha (%)	
Whole period		$7.68^a(5)$	$13.04^a(8)$	$5.78^a(2)$	$42.91^a(11)$	$13.86^a(9)$	$5.41^a(4)$	$26.59^a(7)$	$12.70^a(8)$	3.07(0)
Financial crisis period		$34.47^a(3)$	$34.55^a(1)$	-1.55(0)	$103.24^a(12)$	$37.03^a(9)$	-1.54(0)	$81.20^a(8)$	$33.21^a(8)$	-4.54(0)

Table 8. Price discovery of implied volatilities

This table reports the price discovery measures of implied volatilities. The VECM models are used to calculate the price discovery measures, including the Hasbrouck measure and Gonzalo-Granger measure. The left panel reports the results of the cointegration relationship. The middle panel reports the results of the Hasbrouck measure, while the right panel reports the results of the Gonzalo-Granger measure. In the middle and right panel, the number in maturity row i and maturity column j is the price discovery measure of i-day implied volatility when the VECM model is run between the i-day implied volatility and the j-day implied volatility. The mean price discovery measure of one maturity is the mean of its four individual price discovery measures calculated from the VECM models run against the other maturities.

	Maturity	Cointegration relationship					Hasbrouck measure						Gonzalo-Granger measure					
	(days)	30	91	152	365	730	30	91	152	365	730	Mean	30	91	152	365	730	Mean
ATM call	30		Yes	Yes	Yes	Yes		0.44	0.42	0.40	0.40	0.41		0.10	0.07	0.01	0.07	0.06
	91	Yes		Yes	Yes	Yes	0.56		0.46	0.44	0.46	0.48	0.90		0.25	0.07	0.17	0.35
	152	Yes	Yes		Yes	Yes	0.59	0.54		0.48	0.50	0.53	0.93	0.75		0.08	0.35	0.53
	365	Yes	Yes	Yes		Yes	0.60	0.57	0.53		0.54	0.56	0.99	0.93	0.92		0.85	0.93
	730	Yes	Yes	Yes	Yes		0.60	0.54	0.50	0.46		0.53	0.93	0.83	0.65	0.15		0.64
ATM put	30		Yes	Yes	Yes	Yes		0.44	0.40	0.36	0.43	0.41		0.16	0.14	0.07	0.10	0.12
	91	Yes		Yes	Yes	Yes	0.56		0.44	0.45	0.47	0.48	0.84		0.29	0.03	0.21	0.34
	152	Yes	Yes		Yes	Yes	0.60	0.56		0.48	0.50	0.53	0.86	0.71		0.14	0.37	0.52
	365	Yes	Yes	Yes		Yes	0.64	0.55	0.52		0.55	0.56	0.93	0.97	0.86		0.91	0.92
	730	Yes	Yes	Yes	Yes		0.57	0.53	0.50	0.45		0.51	0.90	0.79	0.63	0.09		0.61
Call with delta = 0.60	30		Yes	Yes	Yes	Yes		0.43	0.41	0.40	0.36	0.40		0.16	0.13	0.03	0.02	0.08
	91	Yes		Yes	Yes	Yes	0.57		0.46	0.45	0.44	0.48	0.84		0.26	0.03	0.14	0.32
	152	Yes	Yes		Yes	Yes	0.59	0.54		0.47	0.48	0.52	0.87	0.74		0.04	0.27	0.48
	365	Yes	Yes	Yes		Yes	0.60	0.55	0.53		0.50	0.55	0.97	0.97	0.96		0.44	0.84
	730	Yes	Yes	Yes	Yes		0.64	0.56	0.52	0.50		0.56	0.98	0.86	0.73	0.56		0.79
Call with delta = 0.40	30		Yes	Yes	Yes	Yes		0.42	0.43	0.42	0.40	0.41		0.15	0.02	0.01	0.07	0.06
	91	Yes		Yes	Yes	Yes	0.58		0.47	0.45	0.49	0.50	0.85		0.14	0.01	0.27	0.32
	152	Yes	Yes		Yes	Yes	0.57	0.53		0.46	0.50	0.52	0.98	0.86		0.04	0.37	0.56
	365	Yes	Yes	Yes		Yes	0.60	0.55	0.54		0.56	0.56	0.99	0.99	0.96		0.98	0.98
	730	Yes	Yes	Yes	Yes		0.60	0.51	0.50	0.44		0.51	0.93	0.73	0.63	0.02		0.58

Table 9. Trading summary of options

This table reports the trading volume and open interest of options from 1996 to 2011. The option data with negative bid-ask spread, negative trading volume and open interest or negative implied volatility are excluded. The trading volume and open interest of ATM (call and put) options, call options with delta equal to 0.60, and call options with delta equal to 0.40 are calculated from the options with moneyness between 45% and 55%, between 55% and 65% and between 35% and 45%, respectively.

Maturity	Volume	Percentage	Open Interest	Percentage	Volume	Percentage	Open Interest	Percentage
(days)	ATM call				ATM put			
<30	24,405,910	27.44%	125,145,667	19.20%	26,889,473	30.24%	127,842,317	19.62%
30∼91	45,275,355	50.91%	226,551,859	34.76%	47,203,201	53.08%	234,450,773	35.98%
91∼152	9,792,743	11.01%	86,922,954	13.34%	10,043,556	11.29%	94,378,002	14.48%
152∼365	6,645,172	7.47%	148,688,840	22.82%	5,375,383	6.04%	125,186,029	19.21%
365∼730	2,188,425	2.46%	53,928,946	8.28%	1,212,805	1.36%	37,068,563	5.69%
>730	619,937	0.70%	10,445,306	1.60%	171,767	0.19%	4,000,570	0.61%
All	88,927,542	100.00%	651,683,572	100.00%	90,896,185	100.00%	622,926,254	100.00%
(days)	Call with delta = 0.60				Call with delta = 0.40			
<30	14,458,604	33.60%	127,079,207	22.87%	20,857,077	40.82%	124,749,489	21.96%
30∼91	18,944,357	44.03%	209,945,899	37.79%	20,219,537	39.57%	196,710,189	34.63%
91∼152	3,627,656	8.43%	70,285,359	12.65%	3,973,515	7.78%	79,691,417	14.03%
152∼365	3,900,412	9.06%	107,205,115	19.30%	4,444,071	8.70%	125,287,852	22.06%
365∼730	1,670,757	3.88%	35,012,580	6.30%	1,387,214	2.72%	36,641,243	6.45%
>730	426,631	0.99%	6,029,213	1.09%	210,379	0.41%	4,976,429	0.88%
All	43,028,417	100.00%	555,557,373	100.00%	51,091,793	100.00%	568,056,619	100.00%

Figure 1. Implied volatility of selected maturities
This graph plots the time series of implied volatilities of selected maturities, including 30 days, 91 days, 152 days, 365 days and 730 days.

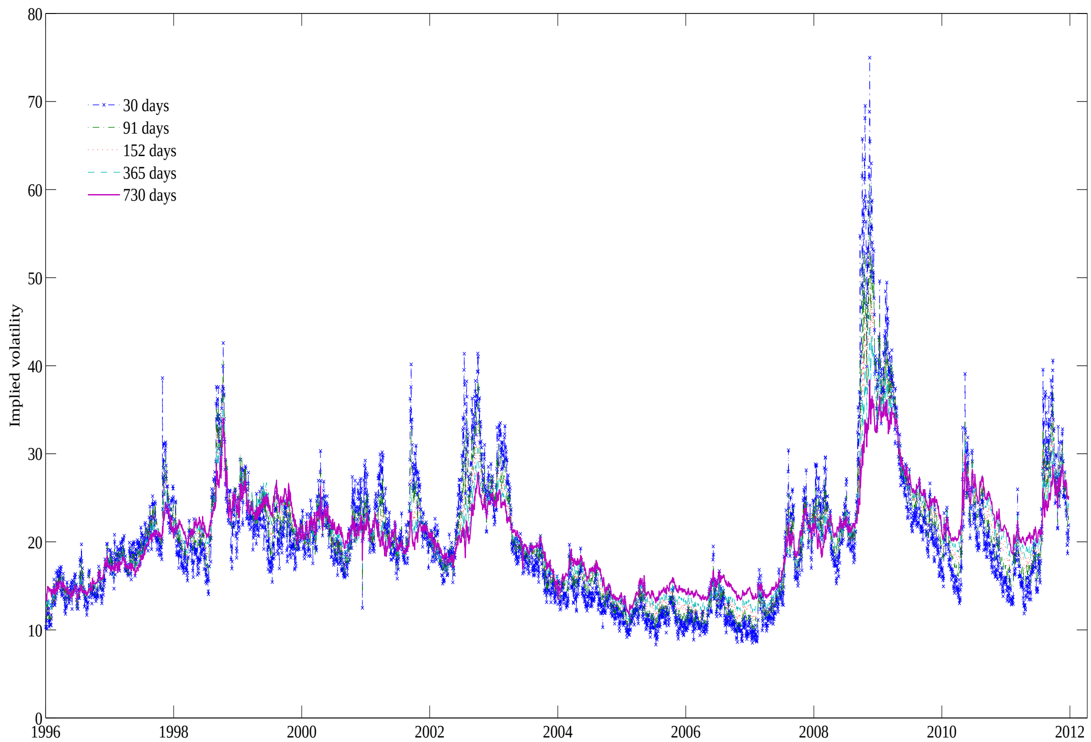

Figure 2. Difference of out-of-sample forecast error between VAR(1) model on volatility change and random walk model

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 random walk model. A negative value means smaller out-of-sample forecast errors for the VAR(1) model on volatility change.

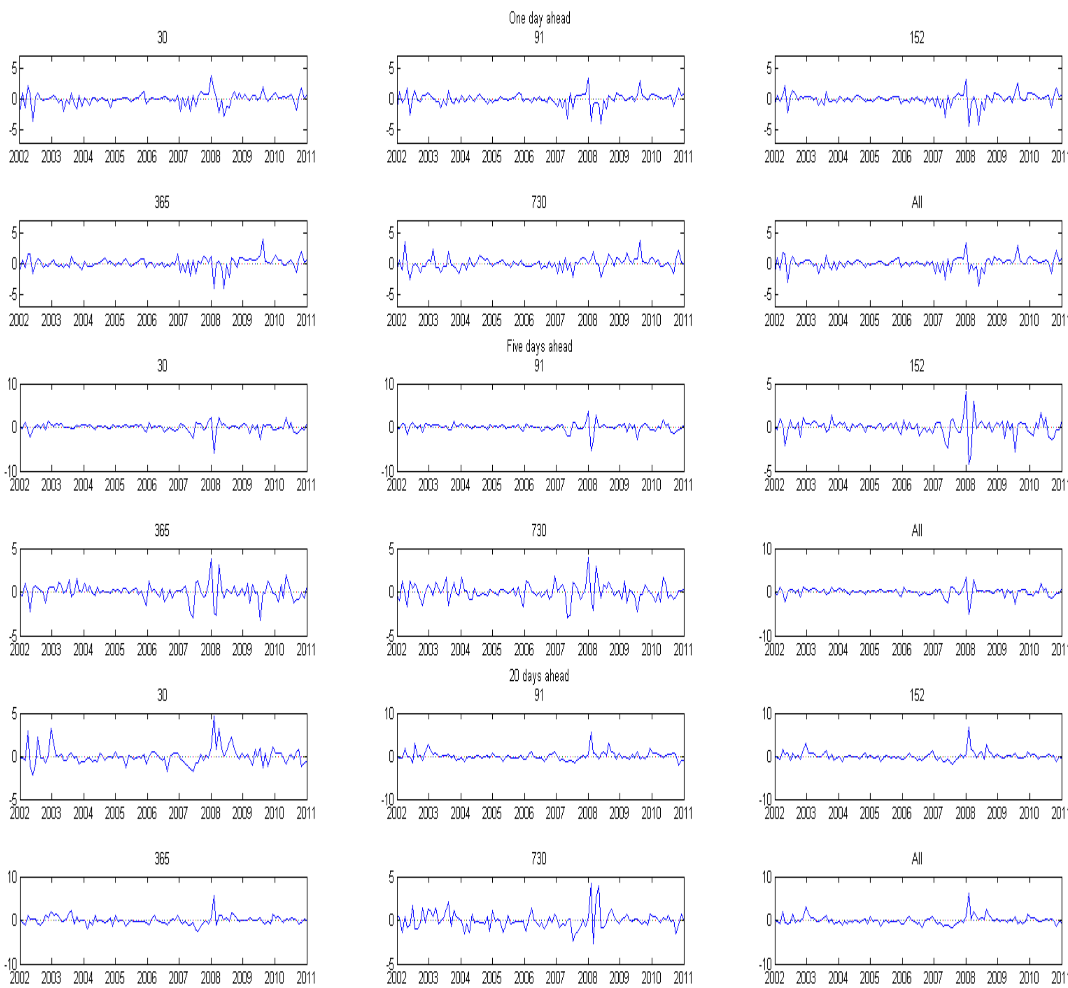

Figure 3. Time series of monthly portfolio return: One-day forecast
This graph plots the monthly return of portfolios that are based on the one-day forecast of implied volatility by 16 different models.

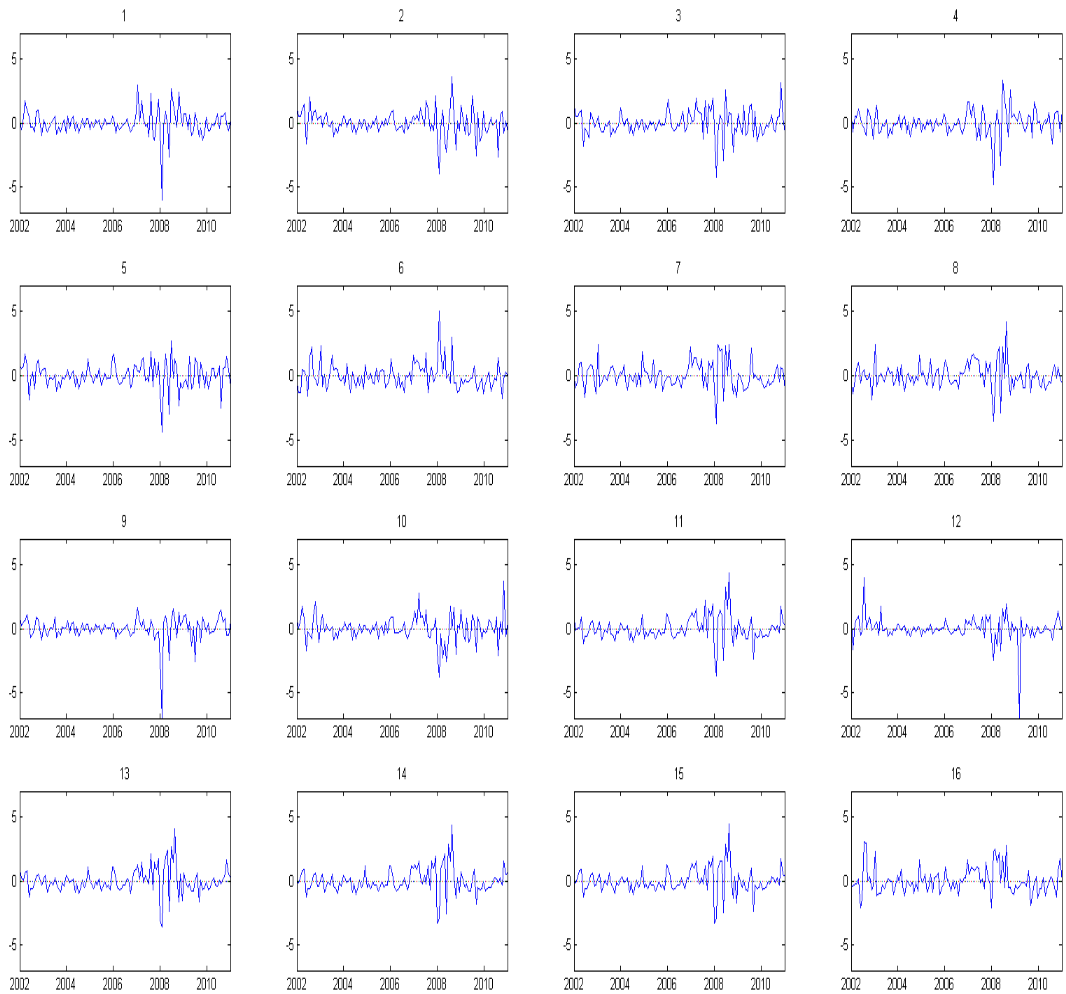

Figure 4. Time series of monthly portfolio return: Five-day forecast
This graph plots the monthly return of portfolios that are based on the five-day forecast of implied volatility by 16 different models.

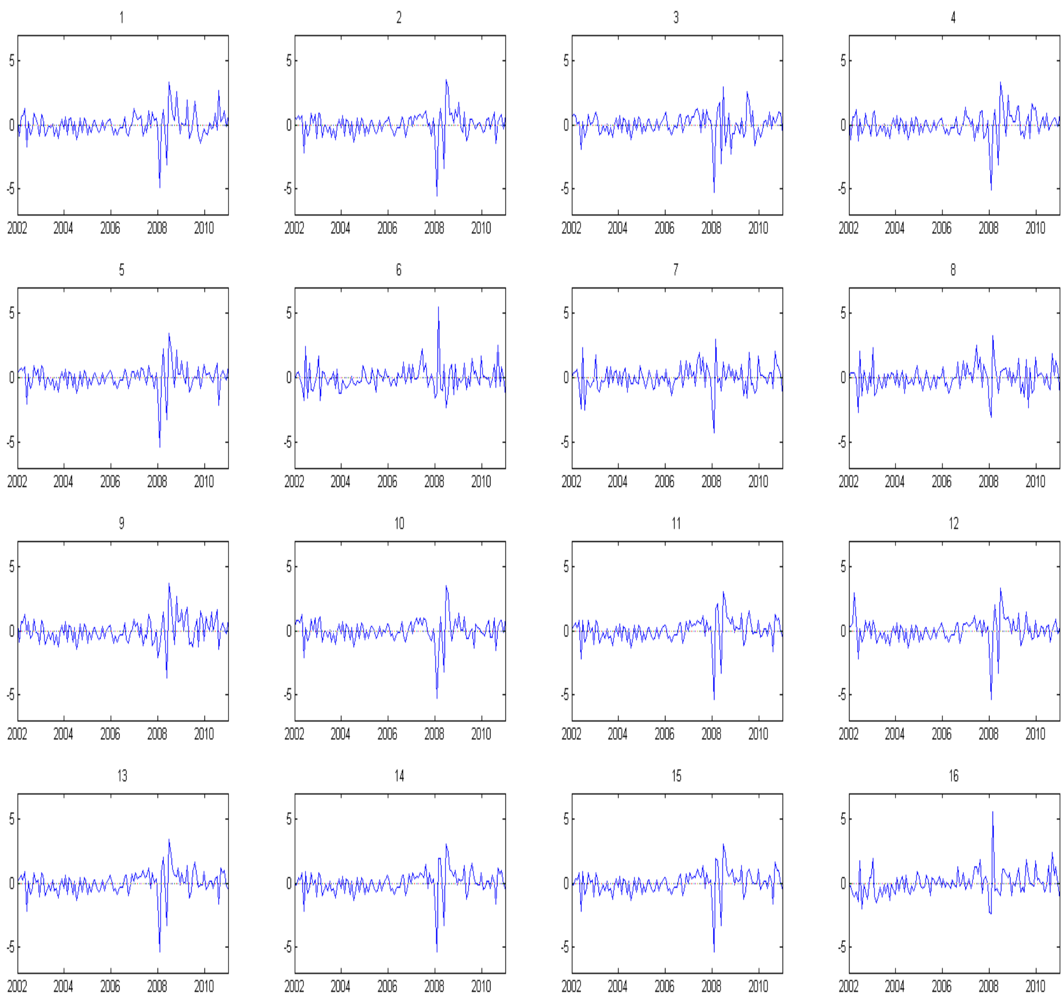

Figure 5. Mean price discovery measure
This graph plots the mean price discovery measure of implied volatilities with different maturities. The mean price discovery measure of one maturity's implied volatility is the mean of its four individual price discovery measures calculated from the VECM models run against the implied volatilities of the other maturities. The up panel plots the results of the Hasbrouck measure, while the bottom panel plots the results of the Gonzalo-Granger measure.