Forecasting the S&P 500 index volatility using investor sentiment

Suk Joon Byun[a] and Hangjun Cho[b,*]

[a]Associate Professor, KAIST Business School, Korea Advanced Institute of Science and Technology, 85 Hoegiro, Dongdaemun-gu, Seoul 130-722, Republic of Korea, Tel: 82-2-958-3352, Fax: 82-2-958-3604, e-mail: sjbyun@business.kaist.ac.kr

[b]Ph.D. Candidate, KAIST Business School, Korea Advanced Institute of Science and Technology, 85 Hoegiro, Dongdaemun-gu, Seoul 130-722, Republic of Korea, Tel: 82-2-958-3968, Fax: 82-2-958-3604, e-mail: chjun01@business.kaist.ac.kr

*Corresponding author, KAIST Business School, Korea Advanced Institute of Science and Technology, 85 Hoegiro, Dongdaemun-gu, Seoul 130-722, Republic of Korea, Tel: 82-2-958-3968, Fax: 82-2-958-3604, e-mail: chjun01@business.kaist.ac.kr

Abstract

We examine several approaches to obtain the volatility forecast for the S&P 500 index: the GARCH-type models, an implied volatility, and their linear combinations. Based on the results, we document that linear combination outperforms the individual models. This result is consistent with existing literature. We also investigate the effect of the regime-switching method using investor sentiment. According to the results, we suggest that the regime-switching method using investor sentiment makes the volatility forecast value more efficient.

EFM classification: 450, 350, 720

1. Introduction

Various forecast models have been applied to forecast the volatility of financial assets. According to Poon and Granger (2003), there are two approaches to volatility forecasting, one that uses time series data and the other that uses option prices. The approach using time series data consists of the historical moving average, the exponentially weighted moving average, stochastic volatility, and the generalized autoregressive conditional heteroskedasticity-type (GARCH-type) models, such as the ARCH and GARCH variation models. This approach uses the past volatility or returns data to forecast future volatility. The option implied volatility (IV) belongs to the other approach that uses option prices. IV is calculated using the Black–Scholes option pricing formula, proposed by Black and Scholes (1973). All the parameters in this formula except for volatility are observed in the financial markets. Therefore, the estimated IV is derived by inverting the formula and using the observed option prices and other parameters.

Although previous studies have investigated a number of volatility forecast models, no consensus has been reached on which model is absolutely better than the other models. Volatility forecast models of all approaches have their own pros and cons (Poon and Granger, 2003). Hence, many researchers have examined the combination of forecast models instead of individual forecast models (Timmermann, 2006; Patton and Sheppard, 2009). Among the combined forecast models, linear combination is the most popular method to estimate the volatility forecast values. The most important issue in this method is choosing the appropriate weights for the individual forecasts in linear combination. The weights of the forecast models can have time-varying as well as constant values. Following Elliott and Timmermann (2005), several approaches following the regime-switching method have been examined to obtain time-varying weights. In this study, we focus on the linear combination of models following the regime-switching method.

Several studies have examined the investor sentiment to find its relationship with the financial markets. Baker and Wurlger (2006, 2007) find that the stock market is related to investor sentiment. In addition, Lee et al. (2002), Han (2008), and Yang and Wu (2011) find that investor sentiment influences the IV or the conditional variance of the stock market. From these studies, we can infer that investor sentiment can be used in volatility forecast models with the regime-switching method. Therefore, we also focus on using investor sentiment in the linear combination with the regime-switching method.

The purpose of this paper is to examine and compare the predictive power of GARCH-type and IV models and their linear combinations. The predictive power of each forecast model is evaluated using four loss functions, the mean square error (MSE), MSE-LOG, mean absolute error (MAE), and MAE-LOG as well as the Diebold–Mariano and the Superior Predictive Ability (SPA) tests. In order to calculate the loss functions more accurately, the realized volatility is estimated based on the intraday data of the S&P 500 index. Furthermore, we investigate the effects of using the regime-switching method for volatility forecasting. For this empirical study, we define two regimes; the determinant of the regimes is the median value of Baker and Wurgler’s (2007) investor sentiment index. We first apply Baker and Wurgler’s (2007) investor sentiment index to the regime-switching method to forecast the volatility of the S&P 500 index.

This paper is organized as follows. Section 2 explains the methodology used in this study and the literature review. Section 3 describes the data used. Section 4 presents the empirical results of the individual forecast models and their linear combinations. Section 5 shows the empirical results of the regime-switching model using investor sentiment. Section

6 presents the results of robustness checks. Section 7 concludes the paper.

2. Methodology and literature review

For individual forecast models, we consider the GARCH-type and IV models. According to Poon and Granger (2003), various GARCH-type models have been used to forecast the volatility of financial assets. The simplest GARCH-type model is the univariate GARCH model of Bollerslev (1986). From this simple GARCH-type model, more complicated GARCH-type models have been investigated. For example, Nelson (1991), Zakoian (1994), and Glosten et al. (1993) propose the Exponential GARCH (EGARCH), the Threshold GARCH (TGARCH), and the GJR-GARCH models, respectively. These are asymmetric volatility models, which have different effects between positive and negative shocks. In this study, we apply a simple GARCH(1,1) model, following Day and Lewis (1992), Brailsford and Faff (1996), Andersen et al. (1999), Yu et al. (2010), and Chuang et al. (2013), and an EGARCH(1,1) model to investigate the existence of asymmetric effect[1]. The specification of a simple GARCH(1,1) model can be written as

$$r_t = c + \varepsilon_t, \qquad \varepsilon_t \sim N(0, h_t) \tag{1}$$

$$h_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}^2, \tag{2}$$

[1] Pagan and Schwert (1990) and Hansen and Lunde (2005) find an asymmetric relation between volatility and past stock returns.

where r_t represents a return series of financial assets. Furthermore, according to Nelson (1991), the specification of an EGARCH(1,1) model can be written as

$$\ln h_t^2 = \omega + \alpha\left(\frac{\varepsilon_{t-1}}{h_{t-1}}\right) + \beta \ln h_{t-1}^2 + \gamma\left[\left|\frac{\varepsilon_{t-1}}{h_{t-1}}\right| - E\left(\left|\frac{\varepsilon_{t-1}}{h_{t-1}}\right|\right)\right], \tag{3}$$

where γ indicates an asymmetric effect. As mentioned earlier, a large number of studies have used GARCH-type models to forecast volatility. Pagan and Schwert (1990) apply the GARCH and EGARCH models to forecast the conditional volatility of monthly stock returns. Day and Lewis (1992) also use the GARCH and EGARCH models to forecast stock market volatility. Brailsford and Faff (1996) compare the GARCH and GJR-GARCH models with other forecasting models to evaluate the accuracy of forecasts for the Australian stock market. In Hansen and Lunde (2005), various GARCH-type models including GARCH and EGARCH are investigated to find a superior forecast model for the exchange rate and stock markets. Koopman et al. (2005) use the GARCH model with the exogenous variable to find the best forecast model for the S&P 100 index volatility. Yu et al. (2010) use the GARCH model to forecast the stock index volatility in Hong Kong and Japan. Chuang et al. (2013) predict the volatilities of the S&P 100 index and equity options using the GARCH model.

IV is widely applied with GARCH-type models to forecast the volatility of financial assets and measures the expected future volatility of an underlying asset. From this point, the primary difference between IV and GARCH-type models is that IV has a forward-looking nature whereas GARCH-type models have backward-looking characteristics, because they are calculated from historical time series data. IV can be obtained from the price of options or the volatility index (VIX) in case of the S&P stock index[2]. Following the instructions for VIX, the S&P 500 index out-of-the-money call and put options prices with the two most near expiration months are used to calculate each IV, and the weighted average of the IVs is quoted as the VIX[3]. According to Poon and Granger (2003), most of the previous studies conclude that IV contains useful information about future volatility of the stock index. Moreover, although no consensus has been reached on which model is the best to forecast volatility, some studies have found the IV model superior to the other forecast methods (see, for example, Fleming et al., 1995; Jorion, 1995; Christensen and Prabhala, 1998; Fleming, 1998; Blair et al., 2001; and Szakmary et al., 2003).

To improve the accuracy of volatility forecasts, a linear combination of individual forecasts is used. When the errors of forecasts are not perfectly correlated and are statistically different, a linear combination of individual forecasts outperforms and holds more information than the individual forecasts. According to De Menezes et al. (2000), Timmermann (2006), and Aiolfi et al. (2011), a large volume of studies have investigated the combination of forecasts. Terui and van Dijk (2002) examine the combination of forecasts to obtain a better forecast model for the US macroeconomic variables. Ang et al. (2007) forecast the US inflation using a composite forecast. Rapach et al. (2010) and Paye (2012) use combined forecasts to predict equity premium and stock market volatility, respectively. In Granger and Ramanathan (1984), three regression equations are considered as linear combinations:

---

[2] Some studies use the VIX as the measure of IV for S&P stock index. See Fleming et al. (1995) and Wang et al. (2006).

[3] For more detailed instructions, see http://www.cboe.com/micro/vix/vixwhite.pdf

$$(i)\; y_{t+1} = \omega_0 + \boldsymbol{\omega}'\hat{\mathbf{y}}_{t+1|t} + \varepsilon_{t+1}, \tag{4}$$

$$(ii)\; y_{t+1} = \boldsymbol{\omega}'\hat{\mathbf{y}}_{t+1|t} + \varepsilon_{t+1}, \tag{5}$$

$$(iii)\; y_{t+1} = \boldsymbol{\omega}'\hat{\mathbf{y}}_{t+1|t} + \varepsilon_{t+1}, \quad s.t.\, \boldsymbol{\omega}'\boldsymbol{\iota} = 1, \tag{6}$$

where y_{t+1} is the ex-post volatility at t+1, $\hat{\mathbf{y}}_{t+1|t}$ is the N-vector of volatility forecasts for t+1 at t, and $\boldsymbol{\iota}$ is an $N \times 1$ vector of ones. Eq. (4) is a standard ordinary least squares (OLS) equation, and Eq. (5) is an OLS equation without an intercept term. When the condition that the sum of weights for forecasts should be one is added, Eq. (5) becomes Eq. (6). According to Benavides and Capistrán (2012), an unbiased composite forecast can be obtained from Eq. (6), but the regression result may not be efficient due to the constraint. Therefore, in this study, we consider Eqs. (4) and (5) as linear combinations of volatility forecasts.

In addition to the simple linear combinations mentioned above, there are several studies on time-varying weights. Following Elliott and Timmermann (2005), there are three methods for time-varying combination weights: use the rolling window method, assume that the weights have their own distribution, and use the regime-switching method. Among these methods, the regime-switching method used in this study indicates a combination of forecasts, with the weights varying depending on the regime. For example, if there are two regimes, a set of weights for one regime is different from that for the other. This example can be expressed as follows:

$$\hat{y}_{t+1|t,c} = I_{(\cdot)}(\alpha_0 + \boldsymbol{\alpha}'\hat{\mathbf{y}}_{t+1|t}) + (1 - I_{(\cdot)})(\beta_0 + \boldsymbol{\beta}'\hat{\mathbf{y}}_{t+1|t}), \tag{7}$$

where $\hat{y}_{t+1|t,c}$ is the combined forecast for t + 1 at t, $I_{(\cdot)}$ is the indicator function for determining the regime, and $(\alpha_0, \boldsymbol{\alpha})$ and $(\beta_0, \boldsymbol{\beta})$ are the respective sets of the weights for two regimes. According to Benavides and Capistrán (2012), how to distinguish the regimes, that is, how to determine when to switch, is the most important question. There are two approaches to distinguish the regimes: one the use of observable variables (see, for example, Deutsch et al., 1994; Bradley and Jansen, 2004), and the other the use of latent variables (see, for example, Elliott and Timmermann, 2005; Ang et al., 2007). Deutsch et al. (1994) consider past forecast errors or macroeconomic variables to construct the indicator function. On the other hand, Ang et al. (2007) use Markov chain probabilities to distinguish the regimes.

Some studies have examined the effect of investor sentiment on the stock market. Baker and Wurgler (2006, 2007) find that investor sentiment influences individual firms and the stock market returns, and Yu and Yuan (2011) show that investor sentiment has an effect on the stock market's mean–variance relation. Moreover, from the study of Han (2008), the S&P 500 index option prices have a relationship with the investor sentiment; hence, investor sentiment can affect the IV of options. With respect to GARCH-type models, Lee et al. (2002) and Yang and Wu (2011) show that investor sentiment has a statistically significant effect on the conditional volatility of the stock market. Therefore, in this study, we propose a regime-switching method using the investor sentiment index of Baker and Wurgler (2007) to determine the regime. In other words, we determine the regime by using the median value of the investor sentiment index: if the value of the investor sentiment index is higher (smaller) than the median, we assgin regime I (regime II). This specification is described as follows:

$$\hat{y}_{t+1|t,c} = I_{(S_t \geq median)}(\alpha_0 + \boldsymbol{\alpha}'\hat{\mathbf{y}}_{t+1|t}) + (1 - I_{(S_t \geq median)})(\beta_0 + \boldsymbol{\beta}'\hat{\mathbf{y}}_{t+1|t}), \tag{8}$$

where S_t is the value of the investor sentiment index at *t*. For our empirical study, we use the following equations:

$$y^i_{t+1} = \omega^i_o + \boldsymbol{\omega}^{i\,\prime}\hat{\mathbf{y}}^i_{t+1|t} + \varepsilon^i_{t+1}, \tag{9}$$

$$y^i_{t+1} = \boldsymbol{\omega}^{i\,\prime}\hat{\mathbf{y}}^i_{t+1|t} + \varepsilon^i_{t+1}, \quad i = 1, 2, \tag{10}$$

where i indicates the regime defined by using the investor sentiment index, and $\hat{\mathbf{y}}^i_{t+1|t}$ contains the forecasts from IV, the GARCH-type models, or both. We estimate the parameters in Eqs. (9) and (10) for IV, the GARCH-type models, and their linear combinations. The parameters of forecasts in Eqs. (9) and (10) are estimated recursively using an expanding window method.

The loss functions for assessing the predictive power of volatility forecasts are the MSE, MSE-LOG, MAE, and MAE-LOG. These statistics are employed to compare the performance of various forecast models (see, for example, Hansen and Lunde, 2005; Patton and Sheppard, 2009; and Patton, 2011). When we denote $\hat{y}_{t+1|t}$ as the volatility forecasts from individual or composite models, the loss functions can be shown as follows:

$$MSE = \sum_{t=1}^{N} (y_{t+1} - \hat{y}_{t+1|t})^2 / N, \tag{11}$$

$$MSE - LOG = \sum_{t=1}^{N} (\ln y_{t+1} - \ln \hat{y}_{t+1|t})^2 / N, \tag{12}$$

$$MAE = \sum_{t=1}^{N} \left| y_{t+1} - \hat{y}_{t+1|t} \right| / N, \tag{13}$$

$$MAE - LOG = \sum_{t=1}^{N} \left| \ln y_{t+1} - \ln \hat{y}_{t+1|t} \right| / N. \tag{14}$$

Although the above statistics are used to directly compare the various forecast models, Diebold and Mariano (1995) and Hansen (2005) present respectively the Diebold–Mariano and the SPA test statistics to evaluate the relative predictive accuracy of the forecast models. To begin with, the Diebold–Mariano test uses the loss function differential between forecast models p and q, and the loss function differential is defined as $d_{p,q,t} = L(y_{t+1}, \hat{y}^{p}_{t+1|t}) - L(y_{t+1}, \hat{y}^{q}_{t+1|t})$, where $L(\ )$ indicates a loss function like Eqs. (11) to (14). The null hypothesis is that the forecast models p and q have equal predictive accuracy, and is expressed as $H_o : E[d_{p,q,t}] = 0$. From this hypothesis and the loss function differential, the Diebold–Mariano test statistic is obtained as follows:

$$DM\_T = \frac{\bar{d}_{p,q}}{\left(asy.var(\bar{d}_{p,q})\right)^{1/2}} \overset{a}{\sim} N(0,1), \tag{15}$$

where $\bar{d}_{p,q} = \sum_{t=1}^{N} d_{p,q,t} / N$ and $asy.var(\bar{d}_{p,q})$ is the asymptotic variance of $\bar{d}_{p,q}$. Since this statistic has the form of a *t*-statistic, we reject or accept H_o by comparing the statistic and the critical value. According to Hansen (2005), the SPA test can be applied for multiple comparisons of the other forecast models with the benchmark model, whereas the Diebold–Mariano test can be used to compare two forecast models pairwise. The null hypothesis of the SPA test is $H_o : E[\mathbf{d}_t] \le \mathbf{0}$, where $\mathbf{d}_t = (d_{1,t}, \cdots, d_{m,t})$, $d_{j,t} = L(y_{t+1}, \hat{y}^0_{t+1|t}) - L(y_{t+1}, \hat{y}^j_{t+1|t})$, and $\hat{y}^0_{t+1|t}$ is the benchmark model. The SPA test statistic can be written as

$$SPA\_T = \max\left[\max_k \frac{\bar{d}_k}{\left(asy.var(\bar{d}_k)\right)^{1/2}}, 0 \right], \tag{16}$$

where $\bar{d}_k = \sum_{t=1}^{N} d_{k,t} / N$. We then decide whether H_o is rejected or accepted, using bootstrapping. In the case of accepting H_o, we can suggest that the benchmark model outperforms the other forecast models.

3. Data

For GARCH-type model forecasts, we use the monthly S&P 500 index data obtained from the Bloomberg data services. From this monthly data, the monthly returns are computed as a log return of the prices of two successive months, that is, $r_t = \ln(p_t / p_{t-1})$. The monthly returns data period is from January 1990 to December 2010, with a total 252 observations. The first one-month-ahead forecasts using the GARCH(1,1) and EGARCH(1,1) models are for January 1996, with 72 observations from January 1990 to December 1995. The remaining forecasts using the GARCH(1,1) and EGARCH(1,1) models are calculated with the expanding window method until we obtain the forecasts for January 2011.

We use the Chicago Board of Options Exchange (CBOE) volatility index VIX monthly data from the Bloomberg data services as the IV of the S&P 500 index options. The VIX monthly data set used in this study runs from December 1995 to December 2010. Usually, the VIX is quoted as an annualized volatility over the following 30 days, shown as a percentage. In other words, the quoted VIX value indicates the annualized implied volatility for the next 30 days. Therefore, we divide the values of the VIX monthly data by using the square root of 12 (the number of months in one year), to obtain the implied volatility for the next month.

The ex-post volatility used to compare the accuracy of various forecast models is the realized volatility with the intraday data of the S&P 500 index, which can be obtained from the Chicago Mercantile Exchange (CME), instead of the daily return. The realized volatility can be expressed as:

$$RV = \sqrt{\sum_{j=1}^{1/\Delta}\left(\ln(p_{t+j\Delta} / p_{t+(j-1)\Delta})\right)^2} \tag{17}$$

where Δ is the intraday data interval. Following Andersen and Bollerslev (1998)[4], we use 5-minute interval log returns to construct the daily realized volatility of the S&P 500 index; we fix the value of Δ in Eq. (17) as 5 minutes. The intraday data set contains the prices from 9:35 a.m. Eastern Standard Time (EST) to 4:00 p.m. EST, and the period of this data set is from January 2, 1996, to January 31, 2011. To obtain the monthly realized volatility of the S&P 500 index, we sum the daily realized variance (squared value of realized volatility) in the same month, divide the summed value into the number of business days during that month, multiply that value by 22 (the average number of business days in a month), and then calculate the square root value of the result. The sample size of all one-month-ahead volatility forecasts and the realized volatility is 181 observations. For evaluating the accuracy of forecasts, the out-of-sample period is from January 2003 to January 2011.

For the investor sentiment proxy, we use Baker and Wurgler's (2007) orthogonalized investor sentiment index[5]. Baker and Wurgler (2007) provide the monthly investor sentiment index data. The data period is from December 1995 to December 2010. The choice of the data period is determined by data availability of the realized volatility and sentiment index.

[Please insert Tables 1 and 2 here.]

[4] Andersen and Bollerslev (1998) suggest 5-minute intervals to keep the balance between microstructure noise and accuracy from using higher sampling frequency.

[5] Baker and Wurgler (2007) obtain the orthogonalized investor sentiment index by using the first principal component of six investor sentiment proxies, which are orthogonalized by a set of macroeconomic indicators. The six sentiment proxies are closed-end fund discount, detrended log turnover, number of initial public offerings (IPOs), first-day return on IPOs, dividend premium, and equity share in new issues. A set of macroeconomic indicators include growth in industrial production; real growth in durable, nondurable, and services consumption; growth in employment; and an NBER recession indicator. The sentiment index data are available at http://people.stern.nyu.edu/jwurgler/

Table 1 gives the summary statistics of the volatility forecasts of the IV and GARCH-type models, the realized volatility, and the sentiment index. Most of all, the volatility forecasts from IV show larger mean and median values than the realized volatility. This result indicates that IV is an upward biased forecast, which is consistent with Fleming (1998). Table 2 shows the matrix of cross-correlations between the volatility forecasts and the realized volatility. We find that the volatility forecasts and the realized volatility are highly correlated with each other. Therefore, we can expect high predictive accuracy for volatility forecasts. Furthermore, the correlation coefficient between the realized volatility and IV forecasts is larger than that between the realized volatility and GARCH-type forecasts. From Table 2, we can infer that the volatility forecasts from IV contain much more information on realized volatility than GARCH-type forecasts.

4. Empirical results from individual forecasts and their linear combination

We estimate the GARCH(1,1) and EGARCH(1,1) model specifications with normal distribution for the error terms in Eqs. (2) and (3) to examine the existence of asymmetric effects. In order to estimate the parameters in these specifications, we use the whole sample period, that is, from January 1990 to December 2010. Bollerslev and Wooldridge's (1992) robust standard errors are used in our estimation procedure. The estimation results are shown in Table 3. Except for ω in the conditional variance equation, all the estimated parameters in the GARCH and EGARCH models are statistically significant at least at the 10% level. Similar to Pagan and Schwert (1990) and Hansen and Lunde (2005), the asymmetric term in the EGARCH model is statistically significant at the 10% level. Furthermore, the negative shocks to the S&P 500 index have different impacts on conditional variance compared to positive shocks. However, although the log likelihood value for the EGARCH model is lower than that for the GARCH model, the lower values for the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) come from the EGARCH and GARCH models, respectively. This result supports the findings that the asymmetric effect in conditional variance of the S&P 500 index exists, although the significance of the asymmetric effect would be marginal.

[Please insert Table 3 here.]

Table 4 shows the in-sample estimation results when using Eqs. (4) and (5) for the individual forecasts and their linear combinations. The *t*-statistics shown in Table 4 are obtained by using Newey and West's (1987) standard errors. The sample period for this in-sample estimation is from January 1996 to January 2011. To begin with, all the estimated weights for the individual forecast models are statistically significant at the 1% level. In case of combined forecasts, only IV has statistically significant weights. This result means that IV contains more information content about the realized volatility than GARCH and EGARCH, which is consistent with the high correlation coefficient between IV and the realized volatility, and the relatively low correlation coefficient between the realized volatility and GARCH or EGARCH in Table 2. Meanwhile, the adjusted R^2 values of regressions for the composite forecasts are higher than those for the individual forecasts. Therefore, this result confirms that the linear combinations of individual forecasts outperform the individual forecasts. This result is consistent with Terui and van Dijk (2002) and Rapach et al. (2010). When we compare the results of Panels A and B, we find that the adjusted R^2 values for the combined model in Panel B are higher than those in Panel A, and that the intercept terms of Panel A are close to zero and statistically insignificant. In other words, the constraint in Eq. (5) that the intercept term should be zero works well in regressions for the linear combinations[6]. Along with the results in Table 3, Table 4 shows that the explanatory power of the individual EGARCH model is slightly better than that of the individual GARCH model. This result also underpins the asymmetric effect of the conditional variance of the S&P 500 index. However, following the adjusted R^2 values in Table 4, the linear combination of IV and GARCH seems to be better than that of IV and EGARCH in case of linear combinations.

[Please insert Table 4 here.]

The evaluation results of the forecast models' predictive accuracy are given in Tables 5 and 6. Table 5 presents the value of loss functions and the rank of forecast models. The linear combination of IV and EGARCH has lower MSE and MSE-LOG values than the other forecast models, irrespective of whether Eq. (4) or (5) is used in regression. On the other hand, the linear combination of IV and GARCH has the lowest MAE and MAE-LOG values among the others. Following this result, we can conclude that a linear combination including the EGARCH model forecasts better than one including the GARCH model when large

[6] This result is consistent with Benavides and Capistrán (2012), who present volatility forecasts results using exchange rate data. They find that the weights for IV and GARCH models are respectively positive and negative. Furthermore, they state that this result is similar to the logic of portfolio formation, and indicate the appropriateness of time-varying weights in the linear combination of forecasts.

shocks to the S&P 500 index price exist. However, with regard to the sum of the models' rank shown in Table 5 and the adjusted R^2 rank shown in Table 4, the lowest value in each panel is seen in the linear combination of IV and GARCH. Therefore, IV with the GARCH model outperforms the individual forecasts and other linear combinations. When we compare the linear combinations of IV and GARCH when using Eqs. (4) and (5), the result in Panel B is better than that in Panel A. This is consistent with Table 4.

[Please insert Tables 5 and 6 here.]

Table 6 reports the Diebold–Mariano test statistics and the SPA test *p*-values for the forecast models. For both tests, the benchmark model is the linear combination of IV and GARCH using Eq. (5). The Diebold–Mariano test statistics have almost negative values, and the statistics for the individual GARCH and EGARCH models are statistically significant at the 1% level. This is similar to the result shown in Table 5. The IV model and the linear combination of IV and EGARCH without the intercept term have negative and positive values according to the loss functions. However, the positive values are not statistically significant. Moreover, with regard to the SPA test *p*-values, we cannot reject the null hypothesis that the benchmark model outperforms other forecast models. Therefore, the linear combination of IV and GARCH using Eq. (5) is the best volatility forecast model among the various forecast models.

5. Empirical results from the regime-switching model using investor sentiment

In this section, we examine the forecast results obtained with the regime-switching model. As mentioned earlier, the regime is determined by comparing the value of the investor sentiment index. If the sentiment index of the previous month is above the median of the sentiment index time series, we assign regime I. For the opposite case, we assign regime II. Table 7 shows the in-sample estimation results of Eqs. (9) and (10) for all the forecast models. Except for linear combinations, all the individual forecast models have statistically significant weights. With regard to linear combinations, only the estimated weight for the IV model is statistically significant. Moreover, the intercept terms in Panel A are close to zero and mostly insignificant. These results are consistent with Table 4. Naturally, the estimated weights of the forecast models and the adjusted R^2 values for the individual models as shown in Table 4 are between those for regime I and regime II in Table 7. When we compare the adjusted R^2 values of each regression for the two regimes, the explanatory power of regime II is seen higher than that of regime I. In other words, if the previous month's investor sentiment is low, the volatility forecast models can explain more about the realized volatility. This seems to result from the volatility forecast noise. If investor sentiment is high, the volatility forecasts will contain noise due to an increase in market participants. Therefore, the realized volatility of the successive month will have a different value compared to the volatility forecasts from last month. This statement is consistent with De Long et al. (1990) and Lemmon and Portniaguina (2006). Following De Long et al. (1990), noise traders participate in the financial market along with increased investor sentiment, and the deviations in price from the fundamental value of financial assets are created. Figure 1 shows the realized and the forecasted volatilities from the IV, GARCH, and EGARCH models corresponding to the regimes. When we compare Panels A and B of Figure 1, the IV model forecast of regime II seems to have a nearly constant difference with the realized volatility, while the difference between the IV model forecast and the realized volatility of the regime I does not seem to be constant. Furthermore, with regard to the forecast from the GARCH and EGARCH models, the forecast values in Panel B seem to be more similar in shape to the realized volatility than the forecast values in Panel A. Thus, we need to distinguish the regime using the investor sentiment to obtain the volatility forecast of good performance.

[Please insert Table 7 and Figure 1 here.]

Tables 8 and 9 show the out-of-sample forecast evaluation results. In these tables, two models in Panel C represent the linear combinations without the intercept term shown in Tables 4, 5, and 6. We insert these models to investigate the effect of the regime-switching method. From Table 8, the loss function values of all the forecast models drop when we use the regime-switching method. Furthermore, three of the four best forecast models corresponding to each loss function come from Panel B, the forecast models with the regime-switching method using Eq. (10). With regard to the rank of each model, for MAE and MAE-LOG, the linear combination of IV and GARCH has a low rank value. This result is consistent with Table 5. When we sum the four rank numbers to find the best forecast model, the linear combination of IV and GARCH using Eq. (10) has the lowest value. Moreover, considering the adjusted R^2 values in Table 7, this model is the best volatility forecast model for S&P 500 index volatility. As a result, the regime-switching volatility forecast models using investor sentiment outperform the forecast models in Panel C[7].

[7] This result is consistent with Deutsch et al. (1994) and Elliott and Timmermann (2005). They state that the time-varying combination of

[Please insert Tables 8 and 9 here.]

Table 9 shows the Diebold–Mariano test statistics and the SPA test *p*-values for the regime-switching volatility forecast models. Following the results shown in Table 8, we set the linear combination of IV and GARCH using Eq. (10) as the benchmark model. The Diebold–Mariano test statistics for individual forecast models and two models in Panel C are entirely negative and almost statistically significant. Moreover, the statistics with positive value are not statistically significant. The SPA test results indicate that the null hypothesis that the benchmark model has superior predictive accuracy cannot be rejected. Therefore, following the above results, to obtain the volatility forecast of the S&P 500 index, we should use the linear combination of IV and GARCH with the regime-switching method using the investor sentiment. Figure 2 compares the linear combination of IV and GARCH using Eq. (10) and the realized volatility of the S&P 500 index.

[Please insert Figure 2 here.]

6. Robustness checks

For robustness checks, we employ two approaches. The first approach is the volatility

---

forecast models outperform the simple linear combination models.

forecast method proposed by Benavides and Capistrán (2012). Benavides and Capistrán (2012) also use the regime-switching method to construct the forecast value for exchange rate volatility. However, the determinant of the regime is the loss differential between two individual forecast models. They forecast the future loss differential based on the historical loss differential data and the past realized volatility data and assign the regimes corresponding to the predicted future loss differential. Following Benavides and Capistrán (2012), we obtain the volatility forecast value using IV and GARCH (or EGARCH). The other approach is the method used by Yu and Yuan (2011). Yu and Yuan (2011) reveal that the stock market mean–variance relationship is affected by investor sentiment, by using Baker and Wurgler's (2007) index. Furthermore, they execute robustness checks exchanging the sentiment index for macroeconomic variables. These macroeconomic variables are the term premium, default premium, interest rate, dividend-price ratio, and consumption surplus ratio, and all these contain business cycle information. Following Yu and Yuan (2011), we employ the above macroeconomic variables to determine the regimes instead of using the sentiment index data[8]. For both approaches, Eq. (10), which makes the best forecast model in the above results, is used as the specification of the combined forecast.

[Please insert Table 10 here.]

[8] Similar to Yu and Yuan (2011), the term premium is defined as the log return difference between 20-year and 3-month T-bills, default premium is defined as the log return difference between BAA and AAA corporate bonds, interest rate is defined as the one-year T-bill log returns, dividend-price ratio is defined as the ratio of the total dividend to the market price of the S&P 500 index, and, finally, the consumption surplus ratio is approximated by a smoothed average of the previous 40-quarter consumption growth, following Wachter (2006). Since the consumption data is on monthly basis, we assign the same value to the months in same quarter. Data on the term premium, default premium, interest rate, and consumption surplus ratio are obtained from the website of the Federal Reserve at St. Louis. Data on the dividend-price ratio are obtained from the website of Robert J. Shiller. These data are available at http://aida.wss.yale.edu/~shiller/data.htm

The loss function values, the Diebold–Mariano test statistics, and the SPA test results are presented in Table 10. In Table 10, the IV + GARCH model refers to the best forecast model, the linear combination of IV and GARCH using Eq. (10) and the regime-switching method with the sentiment index. This model is the benchmark model for two tests. First, all the alternative models show poorer performance compared to the best forecast model with regard to loss function values. Further, all the Diebold–Mariano test statistics are negative, and several values are statistically significant. Finally, the *p*-values of the SPA test indicate that the IV + GARCH model outperforms the other forecast models. Therefore, our volatility forecast model can predict the S&P 500 index volatility well and outperform the existing forecast methods.

7. Conclusion

In this paper, we investigate various forecast models to find the best volatility forecast model to predict the S&P 500 index volatility. We use the IV, GARCH, and EGARCH models and their linear combinations. The linear combinations comprise the IV and GARCH (or EGARCH) models. To evaluate the predictive accuracy of the various forecast models, we calculate the realized volatility of the S&P 500 index with 5-minute intraday data. From the out-of-sample results obtained by using loss functions such as MSE, MSE-LOG, MAE, and MAE-LOG, as well as the Diebold–Mariano and the SPA tests, the linear combination models seem to outperform the individual forecast models. Among the combined forecast models, the linear combination of IV and GARCH without the intercept term is the best forecast model for predicting the S&P 500 index volatility.

Furthermore, we investigate the effect of employing the regime-switching method using investor sentiment. We determine two regimes based on the value of the previous month's sentiment index. The out-of-sample results indicate that all the forecast models improve in accuracy when they use the regime-switching method. Furthermore, a linear combination forecast with the regime-switching method predicts better than one without the regime-switching method. As a result, the best forecast model for predicting the S&P 500 index volatility is the linear combination of IV and GARCH with the regime-switching method and no intercept term. This model has robust forecast accuracy compared to the other determinants for distinguishing the regimes, such as the forecasted loss function value and macroeconomic variables.

References

Aiolfi, M., Capistrán, C., Timmermann, A., 2011. Forecast combination. In: Clements, M.P., Hendry, D.F. (Eds.), The Oxford handbook of economic forecasting. Oxford University Press.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885-905.

Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. Journal of Empirical Finance 6, 457-477.

Ang, A., Bekaert, G., Wei, M., 2007. Do macro variables, asset markets, or surveys forecast inflation better? Journal of Monetary Economics 54, 1163-1212.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645-1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of Economic Perspective 21, 129-151.

Benavides, G., Capistrán, C., 2012. Forecasting exchange rate volatility: The superior performance of conditional combinations of time series and option implied forecasts. Journal of Empirical Finance 19, 627-639.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637-654.

Blair, B., Poon, S.-H., Taylor, S.J., 2001. Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns. Journal of Econometrics 105, 5-26.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. Econometric Reviews 11, 143-172.

Bradley, M.D., Jansen, D.W., 2004. Forecasting with a nonlinear dynamic model of stock returns and industrial production. International Journal of Forecasting 20, 321-342.

Brailsford, T.J., Faff, R.W., 1996. An evaluation of volatility forecasting techniques. Journal of Banking & Finance 20, 419-438.

Christensen, B.J., Prabhala, N.R., 1998. The relation between implied and realized volatility. Journal of Financial Economics 50, 125-150.

Chuang, W.-I., Huang, T.-C., Lin, B.-H., 2013. Predicting volatility using Markov-switching multifractal model: Evidence from S&P 100 index and equity options. North American Journal of Economics and Finance 25, 168-187.

Day, T.E., Lewis, C.M., 1992. Stock market volatility and the information content of stock index options. Journal of Econometrics 52, 267-287.

De Long, J.B., Shleifer, A., Summers, L.G., Waldmann, R.J., 1990. Noise trader risk in financial markets. Journal of Political Economy 98, 703-738.

De Menezes, L.M., Bunn, D.W., Taylor, J.W., 2000. Review of guidelines for the use of combined forecasts. European Journal of Operational Research 120, 190-204.

Deutsch, M., Granger, C.W.J., Teräsvirta, T., 1994. The combination of forecasts using changing weights. International Journal of Forecasting 10, 47-57.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253-263.

Elliott, G., Timmermann, A., 2005. Optimal forecast combination under regime switching. International Economic Review 46, 1081-1102.

Fleming, J., 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. Journal of Empirical Finance 5, 317-345.

Fleming, J., Ostdiek, B., Whaley, R.E., 1995, Predicting stock market volatility: A new measure. Journal of Futures Market 15, 265-302.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3, 197-204.

Han, B., 2008. Investor sentiment and option prices. Review of Financial Studies 21, 387-414.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365-380.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20, 873-889.

Jorion, P., 1995. Predicting volatility in the foreign exchange market. Journal of Finance 50, 507-528.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance 12, 445-475.

Lee, W.Y., Jiang, C.X., Indro, D.C., 2002. Stock market volatility, excess returns, and the role of investor sentiment. Journal of Banking & Finance 26, 2277-2299.

Lemmon, M., Portniaguina, E., 2006. Consumer confidence and asset prices: Some empirical evidence. Review of Financial Studies 19, 1499-1529.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347-370.

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

Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267-290.

Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246-256.

Patton, A.J., Sheppard, K., 2009. Optimal combinations or realised volatility estimators. International Journal of Forecasting 25, 218-238.

Paye, B.S., 2012. ‘Déjà vol’: Predictive regressions for aggregate stock market volatility using macroeconomic variables. Journal of Financial Economics 106, 527-546.

Poon, S.-H., Granger, C.W.J., 2003. Forecasting volatility in financial markets: A review. Journal of Economic Literature 41, 478-539.

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.

Szakmary, A., Ors, E., Kim, J.K., Davidson III, W.N., 2003. The predictive power of implied volatility: Evidence from 35 futures markets. Journal of Banking & Finance 27, 2151-2175.

Terui, N., van Dijk, H.K., 2002. Combined forecasts from linear and nonlinear time series models. International Journal of Forecasting 18, 421-438.

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. North-Holland, Amsterdam.

Wachter, J.A., 2006. A consumption-based model of the term structure of interest rates. Journal of Financial Economics 79, 365-399.

Wang, Y.H., Keswani, A., Taylor, S.J., 2006. The relationships between sentiment, returns and volatility. International Journal of Forecasting 22, 109-123.

Yang, A.S., Wu, M.-L., 2011. Exploring the relationship between investor sentiment and price volatility. Quantitative Finance 11, 955-965.

Yu, J., Yuan, Y., 2011. Investor sentiment and the mean-variance relation. Journal of Financial Economics 100, 367-381.

Yu, W.W., Lui, E.C.K., Wang, J.W., 2010. The predictive power of the implied volatility of options traded OTC and on exchanges. Journal of Banking & Finance 34, 1-11.

Zakoian, J.M., 1994. Threshold heteroskedasticity models. Journal of Economic Dynamics and Control 18, 931-944.

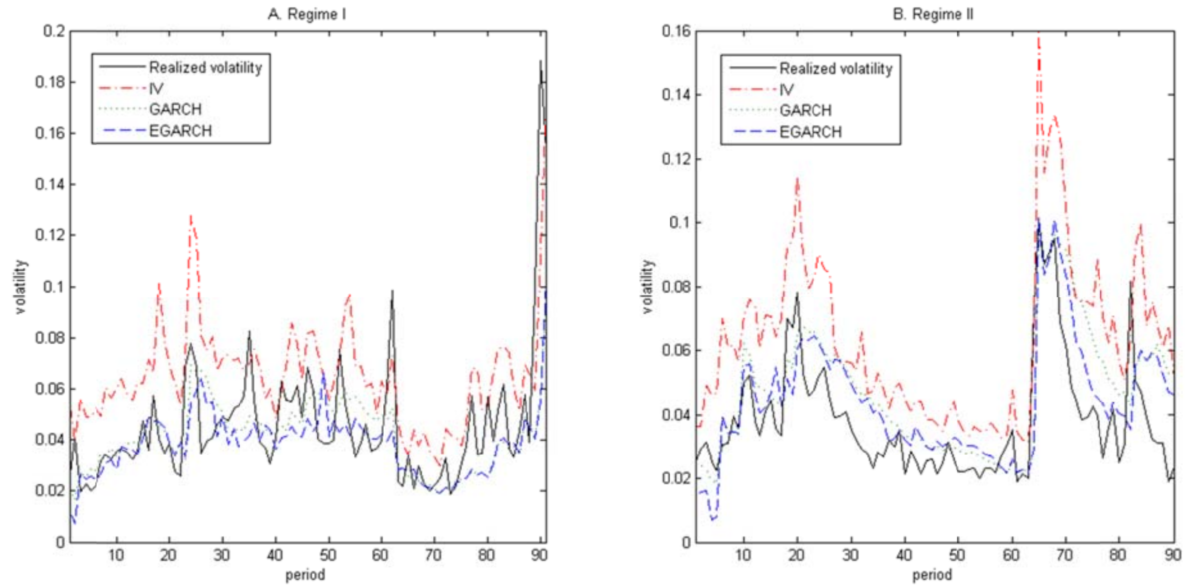

Figure 1 The realized volatility and forecasted volatility from the IV, GARCH, and EGARCH models corresponding to each regime.

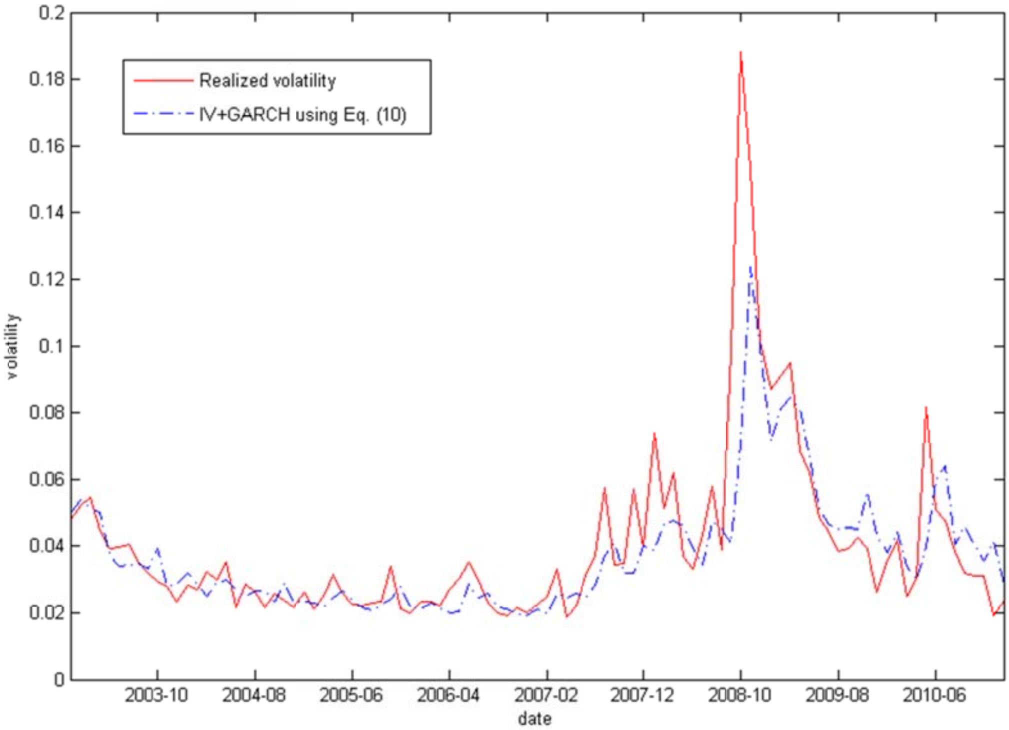

Figure 2 The forecasted value from the linear combination of IV and GARCH using Eq. (10) and the realized volatility of the S&P 500 index.

Table 1

Summary statistics of the volatility forecasts from IV and GARCH-type models, the realized volatility, and the sentiment index

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Obs.
IV forecast	0.0641	0.0620	0.1729	0.0301	0.0236	1.4601	6.5516	181
GARCH forecast	0.0443	0.0435	0.0993	0.0165	0.0172	0.8359	3.7846	181
EGARCH forecast	0.0411	0.0406	0.1029	0.0070	0.0164	1.1150	5.5252	181
Realized volatility	0.0420	0.0364	0.1885	0.0186	0.0223	2.7994	15.5125	181
Sentiment index	0.1835	0.0548	2.4966	-0.9024	0.5891	1.6067	6.0754	181

Note: IV forecast refers to the forecast value from the implied volatility of the S&P 500 index options for one month from now using the VIX monthly data, GARCH forecast refers to the forecast value from the GARCH(1,1) model for one month from now using monthly log returns of the S&P 500 index, EGARCH forecast refers to the forecast value from the EGARCH(1,1) model for one month from now using monthly log returns of the S&P 500 index, Realized volatility refers to the ex-post volatility using the intraday S&P 500 index data, Sentiment index refers to the Baker and Wurgler's (2007) orthogonalized investor sentiment index, Std. Dev. refers to the standard deviation, and Obs. refers to the number of observations. The sample periods for the IV forecast, GARCH forecast, EGARCH forecast, and Sentiment index are from December 1995 to December 2010, and that for the Realized volatility are from January 1996 to January 2011.

Table 2

Matrix of cross-correlations between the volatility forecasts and the realized volatility

	IV forecast	GARCH forecast	EGARCH forecast	Realized volatility
IV forecast	1			
GARCH forecast	0.8445	1		
EGARCH forecast	0.8380	0.9565	1	
Realized volatility	0.7497	0.5805	0.5851	1

Note: The values refer to the correlation coefficients between two variables.

Table 3
Estimation results of the GARCH(1,1) and the EGARCH(1,1) specifications

	GARCH	EGARCH
Mean equation		
c	0.00082***	0.0106***
	(4.5666)	(4.8292)
Conditional variance equation		
ω	9.1602×10^{-5}	-0.3575
	(1.6421)	(-1.5354)
α	0.2703***	0.4128***
	(3.3237)	(3.2027)
β	0.7052***	0.9426***
	(7.8874)	(27.4064)
γ		0.1060*
		(1.7597)
AIC	-903.499	-905.552
BIC	-889.382	-887.905
Log Likelihood	455.750	457.776

Note: *** indicates significance at 1%, and * at 10%. The *t*-statistics are in parentheses below corresponding parameter estimates. Mean equation is Eq. (1), and conditional variance equation is Eq. (2) for the GARCH and Eq. (3) for the EGARCH, respectively.

Table 4

In-sample estimation results of Eq. (4) and (5) for the forecast models

	ω_0	t	ω_1	t	ω_2	t	AIC	BIC	Adj-R^2
Panel A: Eq. (4)									
GARCH	0.0086	1.2699	0.7528***	3.9433			-934.755	-928.358	0.3333
EGARCH	0.0093	1.4874	0.7947***	4.2882			-936.223	-929.826	0.3387
IV	-0.0035	-0.7287	0.7089***	7.4519			-1009.80	-1003.41	0.5596
IV + GARCH	-0.0023	-0.5516	0.8554***	4.7053	-0.2380	-1.4098	-1011.84	-1002.24	0.5669
IV + EGARCH	-0.0027	-0.4857	0.8237***	5.0013	-0.1968	-1.5693	-1010.41	-1000.81	0.5634
Panel B: Eq. (5)									
GARCH			0.9218***	12.5695			-931.499	-928.300	0.3174
EGARCH			0.9893***	14.8540			-931.802	-928.604	0.3186
IV			0.6613***	18.1513			-1010.62	-1007.42	0.5592
IV + GARCH			0.8341***	4.9952	-0.2532	-1.2536	-1013.32	-1006.92	0.5680
IV + EGARCH			0.7957***	6.0329	-0.2117	-1.3767	-1011.68	-1005.28	0.5641

Note: In Panels A and B, the results come from the regression of Eqs. (4) and (5), respectively. *** indicates significance at 1%. The *t*-statistics correspond to the estimated parameters in their left column. ω_0 indicates the intercept term in Eq. (4), and both ω_1 and ω_2 refer to the weights of forecast models. In case of the linear combination, ω_1 and ω_2 correspond to the IV and GARCH-type forecasts, respectively. Adj-R^2 indicates the adjusted R^2 of the regression.

Table 5
MSE, MSE-LOG, MAE, and MAE-LOG from the forecast models

	MSE		MSE-LOG		MAE		MAE-LOG	
	Value	Overall rank	Value	Overall rank	Value	Overall rank	Value	Overall rank
Panel A: Eq. (4)								
GARCH	4.8381×10^{-4}	9	0.1488	8	1.3532×10^{-2}	9	0.3072	8
EGARCH	4.5678×10^{-4}	8	0.1538	9	1.3604×10^{-2}	10	0.3219	10
IV	2.8891×10^{-4}	5	0.0785	6	9.3067×10^{-3}	6	0.2086	6
IV + GARCH	2.8932×10^{-4}	6	0.0771	4	9.1016×10^{-3}	3	0.2044	3
IV + EGARCH	2.8513×10^{-4}	4	0.0764	3	9.1269×10^{-3}	4	0.2048	4
Panel B: Eq. (5)								
GARCH	4.8796×10^{-4}	10	0.1574	10	1.3437×10^{-2}	8	0.3016	9
EGARCH	4.4083×10^{-4}	7	0.1437	7	1.3144×10^{-2}	7	0.3005	7
IV	2.8334×10^{-4}	2	0.0778	5	9.2236×10^{-3}	5	0.2076	5
IV + GARCH	2.8469×10^{-4}	3	0.0761	2	9.0401×10^{-3}	1	0.2032	1
IV + EGARCH	2.7991×10^{-4}	1	0.0754	1	9.0876×10^{-3}	2	0.2042	2

Note: For each criterion, the first column shows the value of the criterion, and the second column shows the models' rank across the forecasts models.

Table 6
The Diebold–Mariano test statistics and the SPA test *p*-values for the forecast models

	MSE	MSE-LOG	MAE	MAE-LOG
Panel A: Eq. (4)				
GARCH	-2.9371***	-5.3901***	-5.6150***	-6.6231***
EGARCH	-2.7844***	-6.1096***	-6.1894***	-6.5181***
IV	-0.8005	-1.0688	-1.1949	-1.0900
IV + GARCH	-0.9111	-1.5294	-0.5037	-0.8108
IV + EGARCH	-0.1041	-0.1465	-0.4960	-0.4025
Panel B: Eq. (5)				
GARCH	-3.4861***	-5.5560***	-5.2414***	-5.7727***
EGARCH	-3.4135***	-5.6197***	-5.8635***	-6.2589***
IV	0.2138	-0.7253	-0.9202	-0.8705
IV + GARCH	-	-	-	-
IV + EGARCH	0.8758	0.3936	-0.3420	-0.2708
SPA_u	0.9286	0.9759	0.9477	0.9541
SPA_c	0.8437	0.8507	0.9095	0.8839
SPA_l	0.4776	0.6447	0.7921	0.7883

Note: For each criterion, column shows the value of the Diebold–Mariano test statistics. *** indicates significance at 1%. The benchmark model for both tests is the linear combination of IV and GARCH using Eq. (5). In other words, the null hypotheses of these Diebold-Mariano statistics are that each model and the linear combination of IV and GARCH without the intercept term have equal predictive accuracy. And, the null hypothesis of the SPA test is that and the linear combination of IV and GARCH without the intercept term outperforms other forecast models.

Table 7

In-sample estimation results of Eqs. (9) and (10) for the forecast models with regime-switching method

	ω_0	t	ω_1	t	ω_2	t	AIC	BIC	Adj-R^2
Panel A: Eq. (9)									
GARCH	0.0021	0.1515	1.0396**	2.6229			-445.231	-440.209	0.3243
	0.0063**	2.2249	0.6833***	7.5672			-530.463	-525.463	0.5301
EGARCH	0.0016	0.1264	1.1452***	2.9619			-446.988	-441.966	0.3372
	0.0066**	2.0200	0.7236***	8.7972			-535.141	-530.141	0.5539
IV	-0.0091	-0.9435	0.8398***	4.1968			-469.700	-464.678	0.4836
	-0.0004	-0.1908	0.6147***	12.9264			-589.762	-584.762	0.7569
IV + GARCH	-0.0092	-0.9063	0.8294***	3.6291	0.0190	0.0943	-467.707	-460.174	0.4778
	0.0005	0.2308	0.7432***	7.2139	-0.1913*	-1.6669	-591.123	-583.624	0.7631
IV + EGARCH	-0.0097	-0.8896	0.7900***	5.0217	0.0993	0.5140	-467.855	-460.323	0.4786
	-0.0000	-0.0200	0.6900***	6.1769	-0.1164	-0.8868	-588.922	-581.423	0.7572
Panel B: Eq. (10)									
GARCH			1.0859***	10.1785			-447.131	-444.620	0.3311
			0.7982***	15.7963			-529.086	-526.587	0.5176
EGARCH			1.1828***	14.0329			-448.930	-446.420	0.3442
			0.8500***	20.8347			-533.120	-530.621	0.5388
IV			0.7130***	9.7010			-469.531	-467.020	0.4770
			0.6091***	24.0436			-591.734	-589.234	0.7595
IV + GARCH			0.7519**	2.3380	-0.0616	-0.1615	-467.602	-462.580	0.4716
			0.7468***	7.5102	-0.1877	-1.5401	-593.089	-588.090	0.7657
IV + EGARCH			0.7067***	4.5392	0.0109	0.0645	-467.533	-462.511	0.4712
			0.6895***	6.6583	-0.1167	-0.8868	-590.922	-585.922	0.7600

Note: In Panels A and B, the results come from the regression of Eqs. (9) and (10), respectively. For each model, estimated result of the regime I (II) is in the upper (lower) row. *** indicates significance at 1%, ** at 5%, and * at 10%. The t-statistics correspond to the estimated parameters in their left column. ω_0 indicates the intercept term in Eq. (4), and both ω_1 and ω_2 refer to the weights of forecast models. In case of the linear combination, ω_1 and ω_2 correspond to the IV and GARCH-type forecasts, respectively. Adj-R^2 indicates the adjusted R^2 of the regression.

Table 8

MSE, MSE-LOG, MAE, and MAE-LOG from the GARCH, EGARCH, and IV models with regime-switching method

	MSE		MSE-LOG		MAE		MAE-LOG	
	Value	Overall rank	Value	Overall rank	Value	Overall rank	Value	Overall rank
Panel A: Eq. (9)								
GARCH	4.1048×10^{-4}	12	0.1173	11	1.1762×10^{-2}	12	0.2618	10
EGARCH	3.7478×10^{-4}	10	0.1166	10	1.1507×10^{-2}	10	0.2698	12
IV	2.6438×10^{-4}	3	0.0698	5	8.3415×10^{-3}	4	0.1894	5
IV + GARCH	2.7520×10^{-4}	6	0.0700	6	8.3362×10^{-3}	3	0.1889	2
IV + EGARCH	2.6599×10^{-4}	4	0.0694	4	8.3326×10^{-3}	1	0.1890	3
Panel B: Eq. (10)								
GARCH	3.9767×10^{-4}	11	0.1235	12	1.1567×10^{-2}	11	0.2621	11
EGARCH	3.4171×10^{-4}	9	0.1056	9	1.0429×10^{-2}	9	0.2403	9
IV	2.5730×10^{-4}	1	0.0685	2	8.3592×10^{-3}	5	0.1898	6
IV + GARCH	2.7172×10^{-4}	5	0.0690	3	8.3353×10^{-3}	2	0.1879	1
IV + EGARCH	2.6250×10^{-4}	2	0.0682	1	8.3670×10^{-3}	6	0.1892	4
Panel C								
IV + GARCH	2.8469×10^{-4}	8	0.0761	8	9.0401×10^{-3}	7	0.2032	7
IV + EGARCH	2.7991×10^{-4}	7	0.0754	7	9.0876×10^{-3}	8	0.2042	8

Note: For each criterion, the first column shows the value of the criterion, and the second column shows the models' rank across the forecasts models. For comparison, two models in Panel C are forecast models without regime-switching method and the intercept term.

Table 9
The Diebold-Mariano test statistics and the SPA test *p*-values for the forecast models with regime-switching method

	MSE	MSE-LOG	MAE	MAE-LOG
Panel A: Eq. (9)				
GARCH	-2.7405***	-4.4204***	-4.9749***	-5.0837***
EGARCH	-2.5767**	-4.4675***	-5.0034***	-5.0416***
IV	1.0932	-0.2359	-0.0277	-0.2344
IV + GARCH	-0.3524	-0.6517	-0.0066	-0.3303
IV + EGARCH	0.9071	-0.1559	0.0133	-0.2000
Panel B: Eq. (10)				
GARCH	-3.1930***	-4.5439***	-4.8892***	-4.9475***
EGARCH	-2.6573***	-3.4176***	-3.7333***	-3.7476***
IV	0.9965	0.1765	-0.1494	-0.4717
IV + GARCH	-	-	-	-
IV + EGARCH	1.0495	0.3807	-0.2590	-0.4145
Panel C				
IV + GARCH	-1.7818*	-2.5269**	-2.6388***	-2.4203**
IV + EGARCH	-0.7501	-1.7883*	-2.8492***	-2.7217***
SPA_u	0.8673	0.9837	0.9652	0.9836
SPA_c	0.4516	0.9078	0.8942	0.9158
SPA_l	0.3041	0.6479	0.8883	0.8541

Note: For each criterion, the column shows the value of the Diebold-Mariano test statistics. *** indicates significance at 1%, ** at 5%, and * at 10%. The benchmark model for both tests is the linear combination of IV and GARCH using Eq. (10). In other words, the null hypotheses of these Diebold-Mariano statistics are that each model and the linear combination of IV and GARCH without the intercept term have equal predictive accuracy. And, the null hypothesis of the SPA test is that and the linear combination of IV and GARCH without the intercept term outperforms other forecast models.

Table 10

MSE, MSE-LOG, MAE, MAE-LOG, the Diebold-Mariano test statistics, and the SPA test *p*-values for robustness checks

	MSE		MSE-LOG		MAE		MAE-LOG	
IV + GARCH	2.7172×10^{-4}	-	0.0690	-	8.3353×10^{-3}	-	0.1879	-
BC_GARCH	4.0151×10^{-4}	-0.9360	0.1014	-2.5896**	1.1789×10^{-2}	-2.7176***	0.2421	-3.6940***
BC_EGARCH	3.3246×10^{-4}	-0.5833	0.0965	-2.1186**	1.0649×10^{-2}	-2.1746**	0.2253	-2.5638**
Term	2.9175×10^{-4}	-2.9811***	0.0788	-3.0761***	9.1782×10^{-3}	-2.8843***	0.2056	-2.5856**
Default	2.9032×10^{-4}	-1.8355*	0.0816	-3.6359***	9.5028×10^{-3}	-3.4716***	0.2132	-3.8208***
Rate	2.9188×10^{-4}	-2.9654***	0.0805	-4.0735***	9.4844×10^{-3}	-4.1232***	0.2137	-4.3253***
D/P	2.7431×10^{-4}	-0.2077	0.0737	-1.3880	8.7395×10^{-3}	-1.2585	0.1970	-1.4233
Surplus	2.8079×10^{-4}	-0.4546	0.0720	-1.0997	8.8235×10^{-3}	-1.7381*	0.1982	-1.7679*
SPA_u	0.9890		0.9896		0.9563		0.9758	
SPA_c	0.9890		0.7042		0.7485		0.7524	
SPA_l	0.6261		0.6024		0.6388		0.6924	

Note: IV + GARCH refers to the linear combination of IV and GARCH using Eq. (10) in Table 7. BC_GARCH and BC_EGARCH indicate the forecast models using the GARCH and the EGARCH, proposed by Benavides and Capistrán (2012). Term, Default, Rate, D/P, and Surplus refer to IV+GARCH models with regime-switching method using macroeconomic variables. Their determinants of the regimes are the term premium, default premium, interest rate, dividend-price ratio, and consumption surplus ratio, respectively. For each criterion, the first column shows the loss function values, and the second column shows the Diebold-Mariano test statistics. *** indicates significance at 1%, ** at 5%, and * at 10% in the second column. The null hypotheses of these Diebold-Mariano statistics are that IV + GARCH and other models have equal predictive accuracy. The benchmark model for the SPA test is IV + GARCH. The null hypotheses of the SPA test are that IV + GARCH outperforms the other models.