Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 543–560
international journal of forecasting
www.elsevier.com/locate/ijforecast

Are VIX futures prices predictable? An empirical investigation

Eirini Konstantinidi[a], George Skiadopoulos[a,b,*]

[a] Department of Banking and Financial Management, University of Piraeus, Greece
[b] Financial Options Research Centre, Warwick Business School, University of Warwick, UK

Abstract

This paper investigates whether volatility futures prices per se can be forecasted by studying the fast-growing VIX futures market. To this end, alternative model specifications are employed. Point and interval out-of-sample forecasts are constructed and evaluated under various statistical metrics. Next, the economic significance of the forecasts obtained is also assessed by performing trading strategies. Only weak evidence of statistically predictable patterns in the evolution of volatility futures prices is found. No trading strategy yields economically significant profits. Hence, the hypothesis that the VIX volatility futures market is informationally efficient cannot be rejected.

Keywords: Bootstrapping; Interval forecasts; Market efficiency; Predictability; Performance measures; VIX; Volatility futures

1. Introduction

Volatility derivatives have attracted a considerable amount of attention in past years, since they enable trading and hedging against changes in volatility. Brenner and Galai (1989, 1993) first suggested derivatives written on some measure of volatility that would serve as the underlying asset. Since then, a number of volatility derivatives have been traded in the over-the-counter market. On March 26, 2004, volatility futures on the implied volatility index VIX were introduced by the Chicago Board Options Exchange (CBOE).[1] Volatility futures on a number of other implied volatility indices have also been introduced since then. The liquidity of volatility futures markets is steadily growing, with the VIX futures market being the most liquid one.[2] This paper focuses on the VIX futures market and for

* Corresponding author at: Department of Banking and Financial Management, University of Piraeus, Greece. Tel.: +30 210 4142363; fax: +30 210 4142341.

E-mail addresses: ekonst@webmail.unipi.gr (E. Konstantinidi), gskiado@unipi.gr (G. Skiadopoulos).

[1] VIX is an implied volatility index that tracks the implied volatility of a synthetic option on the S&P 500 with thirty days to maturity.

[2] The CBOE launched the VXD and VXN volatility futures on April 25, 2005, and July 6, 2007, respectively. The VXD and VXN are implied volatility indices that track the implied volatility of a synthetic option on the Dow Jones Industrial Average and the Nasdaq 100, respectively, with a constant time to maturity (thirty days). Regarding the liquidity of volatility futures, on January 2, 2008, the open interest for VIX futures was 55,792 contracts, or

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.11.004

the first time addresses the question of whether VIX futures prices per se can be predicted.[3] The answer to the question of whether or not volatility futures prices can be predicted is of importance to both academics and practitioners, because it contributes to our understanding of whether volatility futures markets are efficient, and helps market participants to develop profitable volatility trading strategies and set successful hedging schemes.

There is already an extensive body of literature that has investigated whether the prices of stock indexes, interest rates, currencies, and commodity futures can be forecasted. The significance of the results has been evaluated using either a statistical or an economic (trading profits) metric. A number of studies have documented a statistically predictable pattern in futures returns. In particular, Bessembinder and Chan (1992) found that the monthly nearest maturity commodity and currency futures returns can be forecasted within-sample in a statistical sense. They concluded that this predictability could be attributed to an asset pricing model with time-varying risk-premia. Similar findings were documented by Miffre (2001a) for the FTSE 100 futures and by Miffre (2001b) for commodity and financial futures.

On the other hand, the empirical evidence on the predictability in futures markets under an economic metric is mixed. For instance, Hartzmark (1987) found that in aggregate, speculators do not earn significant profits in commodity and interest rate futures markets; daily data of all contract maturities were employed. Yoo and Maddala (1991), however, studied commodity and currency futures and found that speculators tend to be profitable; daily data for a number of futures maturities were considered. Similar findings were reported by Kearns and Manners (2004), Kho (1996), Taylor (1992) and Wang (2004). In particular, all of these studies found that economically significant profits can be obtained by employing various trading rules in currency futures markets; daily data were used by Taylor (1992), and weekly data by Kearns and Manners (2004), Kho (1996), and Wang (2004). A number of futures maturities were examined by Kearns and Manners (2004) and Taylor (1992), while Kho (1996) and Wang (2004) focused on the shortest maturity series. Significant profits were also reported by Hartzmark (1991) and Miffre (2002), who examined the commodity and financial futures markets; the latter study focused only on the shortest maturity contracts. Regarding the source of the identified trading profits, Kearns and Manners (2004) and Taylor (1992) attributed them to the inefficiency of the currency futures market. On the other hand, Kho (1996), Miffre (2002), Wang (2004) and Yoo and Maddala (1991) found that the reported profits were not abnormal, and Hartzmark (1991) found that profitability is determined by luck rather than superior forecast ability; hence, the considered markets were efficient á la Jensen (1978).

In contrast to the number of papers devoted to the topic of predictability in the previously mentioned futures markets, the research as to whether there exist predictable patterns in the evolution of volatility futures prices is still at its infancy. The literature on volatility futures has primarily focused on developing pricing models (see e.g. Brenner, Shu, & Zhang, 2008; Dotsis, Psychoyios, & Skiadopoulos, 2007; Grünbichler & Longstaff, 1996; Lin, 2007; Zhang & Zhu, 2006) and assessing their hedging performances (see e.g. Jiang & Oomen, 2001). On the other hand, to the best of our knowledge, Konstantinidi, Skiadopoulos, and Tzagkaraki (2008) is the only related study that has explored the issue of the predictability of volatility futures prices. However, this was done indirectly, and only under a financial measure. The authors developed trading strategies with VIX and VXD volatility futures based on point and interval forecasts which were formed for the corresponding *underlying* implied volatility indices. They found that the Sharpe ratios obtained were not statistically different from zero, and hence the volatility futures markets are efficient.

$1.3 billion in terms of market value; this corresponds to a 59% increase from January 3, 2007, when the trading volume was 2481 contracts or $57 million in terms of market value. On the same date, the open interest of VXD and VXN futures was $19 and $4 million, respectively.

[3] This question is distinct from the question of whether futures markets are efficient in the sense that the futures price is an optimal forecast of the underlying spot price to be realized on the contract expiry date (see e.g. Coppola, 2008; Kellard, Newbold, Rayner, & Ennew, 1999, and references therein; and Nossman & Vilhelmsson, 2009, for a study using VIX futures). In our study, Jensen's (1978) definition of futures market efficiency is adopted: a market is efficient with respect to the information set I_t in the case where it is impossible to make economic profits by trading on the basis of this information set.

This study extends the literature on whether the evolution of volatility futures prices can be forecasted. In contrast to Konstantinidi et al. (2008), we investigate the predictability of the VIX volatility futures prices per se, without searching for predictable patterns in the underlying implied volatility index. This is because predictability in the underlying implied volatility index market does not necessarily imply that volatility futures prices can be predicted, since there may be other factors/information flows that affect volatility futures markets as well. This is analogous to the interest rate derivatives literature, where it is well documented that models which describe the dynamics of the underlying interest rate quite well, cannot account for the properties of the prices of the corresponding interest rate derivative (the "unspanned stochastic volatility problem"; see, e.g., Jarrow, Li, & Zhao, 2007, and references therein). In our case, the relationship between changes in the prices of VIX futures and its underlying index is not known a priori from a theoretical point of view; there is no cost-of-carry relationship in the case of VIX futures, since the underlying index is not a tradable asset. In addition, volatility futures prices may not always be moving in the same direction as the underlying implied volatility index, due to market microstructure effects (see a discussion and similar findings by Bakshi, Cao, & Chen, 2000, who conducted an analysis for call options using intra-day data).

To address our research question, both point and bootstrapped interval out-of-sample forecasts are considered. This is because interval forecasts have been found to be useful for volatility trading purposes; for example, Poon and Pope (2000) found that profitable volatility spread trades can be developed in the S&P 100 and S&P 500 index option markets by constructing certain intervals. Using a number of tests and criteria, we test the statistical significance of the forecasts obtained. In addition, their economic significance is investigated by means of trading strategies. This is the ultimate test for concluding whether or not the recently inaugurated volatility futures market is efficient. To check the robustness of our results, the analysis is performed across various maturity futures series and by employing a number of alternative model specifications. The latter is necessary because the question of predictability is inevitably tested jointly with the assumed forecasting model.

The remainder of this paper is structured as follows. Section 2 describes the data set, and Section 3 presents the forecasting models to be used. Section 4 discusses the results concerning the in-sample performances of the models under consideration. Next, the out-of-sample predictive performances of the various models are evaluated in statistical and economic terms in Sections 5 and 6, respectively. The last section concludes.

2. The data set

Daily settlement prices of CBOE VIX volatility futures and a set of economic variables are used. The sample period under consideration is from March 26, 2004, to March 13, 2008. The subset from March 18, 2005, to March 13, 2008, is used for the out-of-sample evaluation.

VIX futures were listed by the CBOE in March 26, 2004. They are exchange-traded futures contracts on volatility, and may be used to trade and hedge volatility. The underlying asset of these contracts is VIX. The contract size is $1000 times the VIX.[4] On any given day, the CBOE Futures Exchange (CFE) may list for trading up to six near-term serial months and five months on the February quarterly cycle for the VIX futures contract. The VIX futures contracts are cash settled. The final settlement date is the Wednesday that is thirty days prior to the third Friday of the calendar month immediately following the month in which the contract expires.

The VIX futures prices are obtained from the CBOE website. By ranking the data based on their time to expiration, three time series of futures prices are constructed; namely, the shortest, second shortest, and third shortest maturity series. To minimize the impact of noisy data, we roll over to the next maturity contract five trading days before the contract expires

[4] On March 26, 2007, the VIX futures were rescaled in two ways. First, VIX futures were based on the underlying VIX volatility index directly, instead of on the "Increased-Value index" (VBI = 10∗VIX). Second, the multiplier was increased from $100 to $1000. As a result, the traded futures prices were reduced by a factor of 10, but the $ value of the individual contracts did not change. We adjusted the VIX futures series accordingly.

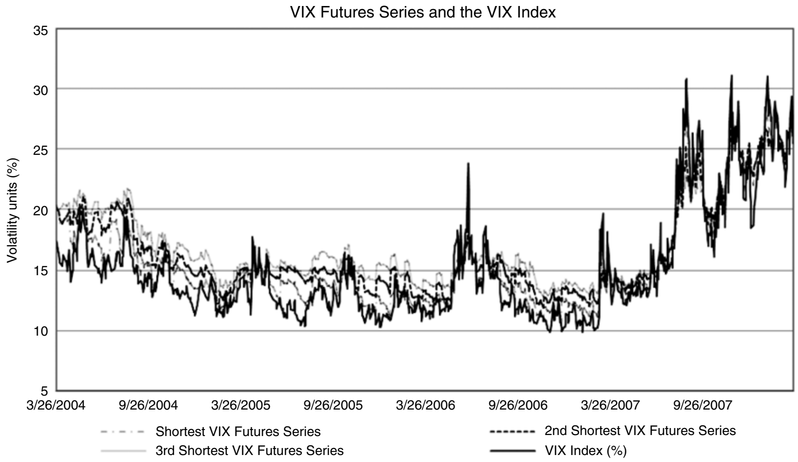

Fig. 1. Evolution of the three shortest maturity VIX futures series and the VIX index over the period March 26, 2004, to March 13, 2008.

(see also Dotsis et al., 2007). Similarly, settlement prices corresponding to a trading volume of fewer than five contracts are excluded.

The data set of economic variables consists of the return on the S&P 500 stock index, the one-month Libor interbank rate, the slope of the yield curve, calculated as the difference between the prices of the ten-year US government bond and the one-month interbank rate, and the basis, calculated as the difference between the VIX index and the VIX futures price for a given maturity T. These data are obtained from Datastream.

Fig. 1 shows the evolution of the three maturity VIX futures series and the VIX index over the period from March 26, 2004, to March 13, 2008. Until March 2007, the term structure of futures prices appears to be upward sloping, with prices being higher for longer maturities (see also Brenner et al., 2008). Table 1 shows the summary statistics for the three series of futures prices and the economic variables in levels and first differences (Panels A and B, respectively). All variables measured in levels are positively (first order) autocorrelated; however, this is not the case when they are measured in first differences. The Augmented Dickey Fuller (ADF, see Dickey & Fuller, 1981) test indicates that most of the VIX futures price series, the S&P 500 and the one-month Libor rate are non-stationary in the levels, but stationary in the first differences. Furthermore, the slope of the yield curve and the basis for most of the futures series are stationary in the levels. The average volume decreases for longer maturities.

3. The forecasting models

3.1. Economic variables model

We use a set of lagged economic variables to forecast the evolution of futures prices. This model specification tests the semi-strong form efficiency of the volatility futures market (see also Bessembinder & Chan, 1992; Kearns & Manners, 2004; Konstantinidi et al., 2008; Miffre, 2001a,b, 2002, for applications of similar predictive specifications to futures markets). Based on the BIC, the following regression is estimated:

$$\Delta F_{t,T} = c + a_1 \Delta F_{t-1,T} + a_2 R_{t-1} + a_3 i_{t-1} + a_4 ys_{t-1} + a_5 basis_{t-1,T} + \varepsilon_{t,T}, \tag{1}$$

where $\Delta F_{t,T}$ denotes the daily changes in the futures price between times $t-1$ and t for a given maturity T ($T = 1, 2, 3$), c is a constant, R_t is the log-return on the S&P 500 stock index between times $t-1$ and t, i_t is the one-month Libor rate in log-differences, ys_t is the slope of the yield curve, and $basis_{t,T}$ is the difference between the VIX index and the VIX futures price for a given maturity T. The variables employed have been shown to have forecasting power in equity markets (see e.g. Welch & Goyal, 2008), and hence they may also have predictive power in

Table 1
Summary statistics. The entries report the summary statistics for each of the VIX futures series and the economic variables. The economic variables under consideration are the S&P 500 stock index, the one-month Libor interbank interest rate, the slope of the yield curve (calculated as the difference between the prices of a ten-year US government bond and the one-month interbank rate), and the basis for each of the VIX futures series. The first order autocorrelation, ρ_1, and the Jarque-Bera and Augmented Dickey Fuller (ADF) test values are also reported. The null hypotheses for the Jarque-Bera and ADF tests are that the series is normally distributed and has a unit root, respectively.

	Shortest	2nd shortest	3rd shortest	S&P 500	1M int. rate	Slope of yield curve	Basis for shortest	Basis for 2nd shortest	Basis for 3rd shortest
Panel A: Summary statistics of VIX futures and economic variables (levels): March 26, 2004, to March 17, 2005									
#Observations	241	235	195	255	250	250	241	235	195
Mean	158.10	169.50	178.56	1143.97	1.83	2.48	−11.99	−24.05	−33.43
Std. deviation	22.44	24.12	24.35	41.20	0.56	0.70	11.48	11.87	11.95
Skewness	0.06	−0.08	−0.22	0.25	0.15	0.26	−0.80	−0.38	0.19
Kurtosis	2.01	1.88	1.85	1.81	1.66	1.67	3.47	2.75	2.17
Jarque-Bera	**10.05***	**12.58***	**12.36***	**17.58***	**19.75***	**21.38***	**27.66***	**6.17****	**6.84****
ρ_1	**0.926***	**0.922***	**0.802***	**0.979***	**0.968***	**0.975***	**0.856***	**0.825***	**0.729***
ADF	**−3.76****	−2.96	−1.66	−2.24	−2.71	**−3.67****	**−4.19***	**−4.70***	−1.21
Mean volume	186.17	135.03	104.16						
(min, max)	(*5–1218*)	(*5–865*)	(*5–974*)						
Panel B: Summary statistics of VIX futures and economic variables (daily differences): Mar 26, 2004, to Mar 17, 2005									
Mean	−0.2689	−0.2498	−0.4430	0.0003	0.0038				
Std. deviation	5.04	4.19	3.30	0.01	0.01				
Skewness	1.86	0.65	0.54	−0.13	1.24				
Kurtosis	11.62	8.45	5.12	2.95	7.34				
Jarque-Bera	**838.58***	**286.46***	**37.41***	0.78	**254.16***				
ρ_1	−0.004	0.082	0.042	0.039	**0.247***				
ADF	**−14.93***	**−13.65***	**−10.85***	**−15.36***	**−5.80***				

* Denotes rejection of the null hypothesis at the 1% level.
** Denotes rejection of the null hypothesis at the 5% level.

futures markets. In addition, the slope of the term structure of interest rates has been shown to be able to forecast forthcoming recessions (see e.g. Estrella & Hardouvelis, 1991), and hence increases in volatility (Schwert, 1989) and volatility futures prices. Fig. 2 shows the evolution of the slope of the yield curve over the period March 26, 2004, to March 13, 2008. Furthermore, the basis may have the ability to forecast the futures risk premium, since it can be decomposed into two terms: the risk premium of the futures contract and the expected change in the underlying asset price (Fama & French, 1987). Finally, notice that changes in the underlying implied volatility index have not been used as an additional predictive variable, in order to avoid multi-collinearity issues; VIX is highly correlated with the VIX futures prices.

3.2. Univariate autoregressive, ARMA and VAR models

Univariate autoregressive, ARMA and VAR models are employed to investigate the extent to which past volatility futures prices can be exploited for predictive purposes, and to examine whether there are spillovers between the three futures series. These model specifications set up tests of weak form market efficiency. The number of lags employed is chosen on the basis of the BIC, and in order to avoid over-fitting the data

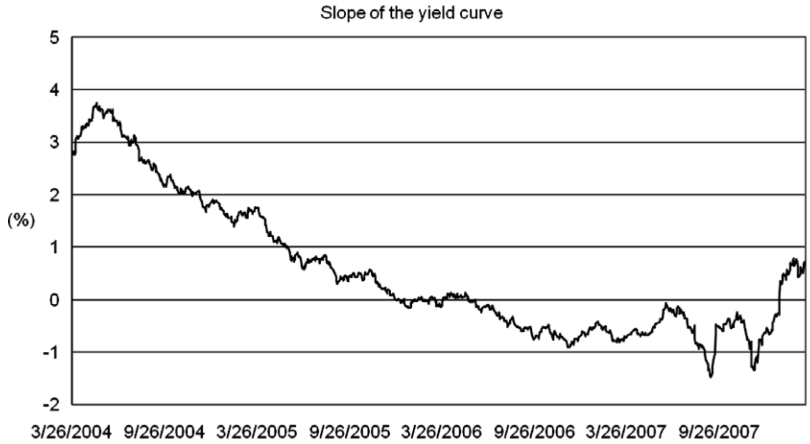

Fig. 2. Evolution of the slope of the yield curve over the period March 26, 2004, to March 13, 2008. The slope of the yield curve is calculated as the difference between the prices of the ten-year US government bond and the one-month interbank rate.

(the maximum number of lags considered was four). The following AR(2) model is estimated:

$$\Delta F_{t,T} = c + \varphi_1 \Delta F_{t-1,T} + \varphi_2 \Delta F_{t-2,T} + \varepsilon_{t,T}. \quad (2)$$

We also estimate an ARMA(1, 1) model:

$$\Delta F_{t,T} = c + \varphi_1 \Delta F_{t-1,T} + \theta_1 \varepsilon_{t-1,T} + \varepsilon_{t,T}. \quad (3)$$

Furthermore, the following VAR(1) model is estimated:

$$\Delta F_t = C + \Phi_1 \Delta F_{t-1} + \varepsilon_t, \quad (4)$$

where ΔF_t is the (3×1) vector of changes in the three futures prices series that are assumed to be jointly determined, C is a (3×1) vector of constants, Φ_1 is a (3×3) matrix of coefficients, and ε_t is a (3×1) vector of residuals.

3.3. Combination forecasts

In addition to model based forecasts, we also consider combination forecasts. Combination forecasts aggregate the information used by the individual forecasting models. They have been found to be more accurate than individual forecasts (see, e.g. Bates & Granger, 1969; and Clemen, 1989, for reviews).

Two alternative linear combination forecasts are considered. First, an equally weighted combination forecast is employed:

$$\hat{F}^{EW}_{t|t-1,T} = \frac{1}{4} \sum_{i=1}^{4} \hat{F}^{i}_{t|t-1,T}, \quad (5)$$

where $\hat{F}^{i}_{t|t-1,T}$ is the forecasted futures price constructed at time $t-1$ for time t for a given maturity T using the ith model specification ($i = 1$ (economic variables model), 2 (AR(2) model), 3 (ARMA(1,1) model), 4 (VAR model)), and $\hat{F}^{EW}_{t|t-1,T}$ denotes the equally weighted combination forecast of the futures price constructed at time $t-1$ for time t for a given maturity T. This is a simple average of all model based forecasts; there is evidence that simple combinations frequently outperform more sophisticated ones (see e.g. Clemen, 1989).

Second, an unequally weighted average of the individual forecasts is used; the weights are chosen so as to minimize the mean squared forecast error (see Granger & Ramanathan, 1984). To illustrate, at time t, the weights are obtained by recursively estimating the following OLS regression:

$$\Delta F_{t,T} = c + \sum_{i=1}^{4} a_i \Delta \hat{F}^{i}_{t|t-1,T} + \varepsilon_{t,T}, \quad (6)$$

where $\Delta F_{t,T}$ is the realized futures price change between times $t-1$ and t for a given maturity T, c is a constant, $\Delta \hat{F}^{i}_{t|t-1,T} = \hat{F}^{i}_{t|t-1,T} - F_{t-1,T}$ is the forecasted futures price change between times $t-1$ and t for a given maturity T, and $\varepsilon_{t,T}$ is the error term. Then:

$$\hat{F}^{W}_{t+1|t,T} = F_{t,T} + c + \sum_{i=1}^{4} a_i \Delta \hat{F}^{i}_{t+1|t,T}, \quad (7)$$

where $\hat{F}^{W}_{t+1|t,T}$ denotes the weighted combination forecast of the futures price constructed at t for time $t+1$ for a given maturity T.

To start recursively constructing the weighted combination forecasts, one needs an "initial" time series of individual forecasts in order to estimate regression (6). To this end, the in-sample data (from March 26, 2004, to March 17, 2005) are divided into an in-sample period (from March 26, 2004, to September 24, 2004) and a "pseudo" out-of-sample period (from September 27, 2004, to March 17, 2005). First, the in-sample data are used to estimate the model specifications described in Sections 3.1–3.2 (Eqs. (1)–(4)). Then, forecasts are formed recursively over the "pseudo" out-of-sample period by adding each observation of the "pseudo" out-of-sample data set to the in-sample data set as it becomes available. Finally, the individual forecasts over the "pseudo" out-of-sample period are used to estimate regression (6). Then, the first out-of-sample weighted combination forecast (corresponding to March 18, 2005) is constructed as described by Eq. (7). To form the remaining out-of-sample combination forecasts, Eq. (6) is estimated recursively by adding each individual forecast to the sample as it becomes available, and Eq. (7) is re-applied.

4. In-sample evidence

Tables 2 and 3 show the in-sample performances of the economic variables and the AR(1)/ARMA(1,1)/VAR models, respectively. The estimated coefficients, the t-statistics (in parentheses), the unadjusted R^2 and the adjusted R^2 are reported for each one of the implied volatility indices. * and ** indicate that the estimated parameters are statistically significant at the 1% and 5% levels, respectively.

In the case of the economic variables model (Table 2), we can see that the adjusted R^2 takes the largest value for the third shortest series (0.9%). This is similar to the values of the adjusted R^2 which have been documented by previous related papers in various futures markets (see Bessembinder & Chan, 1992; Konstantinidi et al., 2008; Miffre, 2001a,b, 2002). In the case of the AR(2) and VAR models (Table 3, Panels A and B, respectively), we can see that the largest values of the adjusted R^2 are obtained for the second shortest series (1.9% and 3.1%, respectively). Finally, the application of the ARMA model (Table 3, Panel C) reveals that there is a predictable pattern in the case of the shortest futures series (the adjusted R^2 is 5.5%).

In summary, the in-sample goodness-of-fit depends on the model specification and the maturity of the futures series under consideration. Next, the out-of-sample performance is assessed in order to provide a firm answer to the question of whether volatility futures prices can be forecasted.

5. Out-of-sample evidence: Statistical significance

Point and bootstrapped interval forecasts are used to assess the out-of-sample performances of the models described in Section 3. The out-of-sample period is from March 18, 2005, to March 13, 2008. To form the point forecasts, the models are initially estimated over the in-sample period (from March 26, 2004, to March 17, 2005), and the first out-of-sample point forecast is obtained (corresponding to March 18, 2005). To construct the remaining out-of-sample point forecasts, the models are re-estimated recursively by adding each observation to the in-sample data set as it becomes available. The bootstrapped interval forecasts are constructed by applying the methodology suggested by Pascual, Romo, and Ruiz (2001), in order to take into account the non-normality of the residuals of the various models and the parameter uncertainty. To this end, 1000 bootstrap samples are formed for each time step (i.e., each day).

5.1. *Point forecasts: Statistical testing*

Three alternative metrics are employed to assess the statistical significance of the out-of-sample point forecasts obtained. The first metric is the root mean squared prediction error (RMSE), calculated as the square root of the average squared deviations of the actual volatility futures prices from the model based forecast, averaged over the number of observations. The second metric is the mean absolute prediction error (MAE), calculated as the average of the absolute differences between the actual volatility futures price and the model based forecast, averaged over the number of observations. The third metric is the mean correct prediction (MCP) of the direction of volatility futures price changes, calculated as the average

Table 2
Forecasting with the economic variables model: in-sample analysis. The entries report results from the regression of each VIX futures series on a set of lagged economic variables, augmented by an AR(1) term. The following specification is estimated: $\Delta F_{t,T} = c + a_1 \Delta F_{t-1,T} + a_2 R_{t-1} + a_3 i_{t-1} + a_4 ys_{t-1} + a_5 basis_{t-1,T} + \varepsilon_{t,T}$, where $\Delta F_{t,T}$: daily changes in the futures prices between times $t-1$ and t for a given maturity T; c: a constant; R_t: the log-return on the S&P 500 stock index between times $t-1$ and t; i_t: the one month Libor rate in log-differences; ys_t: the slope of the yield curve (calculated as the difference between the prices of the ten-year US government bond and the one-month interbank rate); and $basis_{t,T}$: the difference at any time t between the VIX index and the VIX futures price for a given maturity T. The estimated coefficients (with Newey-West t-statistics in parentheses) and the unadjusted and adjusted R^2 are reported. The model has been estimated for the period March 26, 2004, to March 17, 2005.

	Coeff. *(t-stat)*		
	Dependent variable: Shortest	Dependent variable: 2nd shortest	Dependent variable: 3rd shortest
Observations included	212	200	133
C	−0.650	−0.197	−1.070
	(−0.541)	*(−0.173)*	*(−0.886)*
$\Delta F_{t-1,T}$	−0.069	0.052	−0.119
	(−1.225)	(0.840)	*(−1.205)*
R_{t-1}	−76.264	−35.701*	**−134.978****
	(−1.717)	*(−0.881)*	*(−2.323)*
i_{t-1}	14.042	49.234	24.183
	(0.157)	(0.711)	(0.395)
ys_{t-1}	0.425	−0.003	−0.074
	(0.703)	(0.004)	(0.154)
$basis_{t-1,T}$	0.055	0.011	−0.024
	(1.627)	(0.286)	*(−0.736)*
R^2	**0.023**	**0.015**	**0.046**
Adj.R^2	**−0.001**	**−0.011**	**0.009**

* Denotes rejection of the null hypothesis of a zero coefficient at the 1% level.
** Denotes rejection of the null hypothesis of a zero coefficient at the 5% level.

frequency (percentage of observations) for which the change in the volatility futures price predicted by the model has the same sign as the realized change. The forecasts are compared with those obtained from the random walk, which is used as the benchmark model. To this end, we perform pairwise comparisons based on the modified Diebold and Mariano (1995) test (MDM, see Harvey, Leybourne, & Newbold, 1997) and a ratio test for the RMSE/MAE and MCP metrics, respectively. The null hypothesis is that the model under consideration and the random walk perform equally well. Moreover, we use White's (2000) test (also termed a reality check) to *jointly* compare all forecasts to the benchmark model under the RMSE and MAE metrics.[5] In this case, the null hypothesis is that no model outperforms the random walk.

To illustrate, the two tests are described as follows. Let $\left\{\hat{F}^i_{t|t-1,T}\right\}^n_{t=1}$ and $\left\{\hat{F}^{RW}_{t|t-1,T}\right\}^n_{t=1}$ denote the sequence of forecasted futures price from $t = 1, \ldots, n$ based on the ith model ($i = 1$ (economic variables model), 2 (AR(2) model), 3 (VAR model), 4 (ARMA(1, 1) model), 5 (equally weighted combination forecast), 6 (weighted combination forecast)) and the random walk, respectively. Define a loss function $g\left(e^i_{t,T}\right)$ and $g\left(e^{RW}_{t,T}\right)$, and the loss differential $d^i_{t,T} = g\left(e^i_{t,T}\right) - g\left(e^{RW}_{t,T}\right)$, with $\left\{e^i_{t,T}\right\}^n_{t=1}$ and $\left\{e^{RW}_{t,T}\right\}^n_{t=1}$ being the respective forecast errors for the ith model specification and the T-maturity futures series.

[5] Note that the MCP cannot be calculated for the random walk model. However, we proxy the random walk with the naïve rule that "the predicted change in the futures prices has a 50% chance of being positive and a 50% of being negative". This is to say that the random walk case corresponds to an MCP equal to 50%. Similarly, White's (2000) test cannot be applied to the MCP metric. This is because the corresponding loss function cannot be defined for the benchmark model.

Table 3
Forecasting with the univariate autoregressive, ARMA and VAR models: in-sample analysis. Panel A: The entries report results from the estimation of a univariate AR(2) specification for the daily changes of each VIX futures series, namely: $\Delta F_{t,T} = c + \varphi_1 \Delta F_{t-1,T} + \varphi_2 \Delta F_{t-2,T} + \varepsilon_{t,T}$. Panel B: The entries report the estimated coefficients of a VAR, for the three VIX futures prices series: $\Delta F_t = C + \Phi_1 \Delta F_{t-1} + \varepsilon_t$, where ΔF_t is the 3×1 vector of changes in the three futures prices series, C is a (3×1) vector of constants, Φ_1 is the (3×3) matrix of coefficients to be estimated, and ε_t is a (3×1) vector of errors. Panel C: The entries report the estimated coefficients of an ARMA(1, 1) model: $\Delta F_{t,T} = c + \varphi_1 \Delta F_{t-1,T} + \theta_1 \varepsilon_{t-1,T} + \varepsilon_{t,T}$. The estimated coefficients (with Newey-West t-statistics in parentheses), and the unadjusted and adjusted R^2 are reported. The models have been estimated for the period March 26, 2004, to March 17, 2005.

	Coeff. *(t-stat)* Dependent variable: Shortest	Dependent variable: 2nd shortest	Dependent variable: 3rd shortest
Panel A: AR(2) model			
Included obs.	204	191	95
c	−0.465	−0.426	−0.513
	(−1.485)	*(−1.590)*	*(−1.638)*
φ_1	0.034	0.119**	0.021
	(0.746)	(1.986)	(0.280)
φ_2	−0.085	−0.124	−0.089
	(−1.565)	*(−1.542)*	*(−1.048)*
R^2	**0.010**	**0.029**	**0.009**
Adj.R^2	**0.001**	**0.019**	**−0.008**
Panel B: VAR model			
Included obs.	**130**	**130**	**130**
C	−0.524	−0.505	−0.306
	(−1.343)	*(−1.752)*	*(−0.992)*
$\Delta F1_{t-1}$	−0.161	−0.134	−0.121
	(−0.986)	*(−1.113)*	*(−0.941)*
$\Delta F2_{t-1}$	**0.425****	**0.311****	**0.358****
	(1.987)	(1.970)	(2.120)
$\Delta F3_{t-1}$	−0.071	0.005	−0.149
	(−0.289)	(0.028)	*(−0.766)*
R^2	**0.043**	**0.053**	**0.037**
Adj.R^2	**0.020**	**0.031**	**0.014**
Panel C: ARMA(1,1) model			
Included obs.	226	217	156
c	**−0**.042*	−0.373	−0.722
	(−3.875)	*(−0.909)*	*(−1.402)*
φ_1	**0.854***	**(−0.558)**	−0.579
	(25.652)	*(−2.352)*	*(−0.890)*
θ_1	**−0**.992*	**0.679***	0.652
	(−161.116)	*(3.057)*	*(1.062)*
R^2	**0.064**	**0.023**	**0.007**
Adj.R^2	**0.055**	**0.014**	**−0**.006

* Denotes rejection of the null hypothesis of a zero coefficient at the 1% level.
** Denotes rejection of the null hypothesis of a zero coefficient at the 5% level.

In the case of the MDM test, the null hypothesis is $H_0 : E\left(d^i_{t,T}\right) = 0$. We test this against two alternative hypotheses. The first alternative hypothesis is that the random walk outperforms the respective model, i.e. $H_1 : E\left(d^i_{t,T}\right) > 0$. The second alternative hypothesis is that the model under consideration outperforms the random walk, i.e., $H_2 : E\left(d^i_{t,T}\right) < 0$.[6] In the case of one-step-ahead forecasts, the MDM test statistic MDM^i_T for the ith model specification and the T-maturity futures series is given by:

$$MDM^i_T = \frac{\bar{d}^i_T}{\sqrt{\text{var}(\bar{d}^i_T)}}, \tag{8}$$

with $\bar{d}^i_T = \frac{\sum_{t=1}^n d^i_{t,T}}{n}$ and $\text{var}(\bar{d}^i_T)$ being a Newey and West (1987) estimator of the variance of $\bar{d}^i_T$, where Barlett's kernel was employed and the required lag selection parameter was set equal to $[4(n/100)^{2/9}]$. Following Harvey et al. (1997), we make accept/reject decisions by comparing the calculated test statistic to the critical values from the Student's t distribution with $(n-1)$ degrees of freedom.

The idea of White's (2000) test is as follows (see also Sullivan, Timmermann, & White, 1999). At any point in time t, for the ith model and for a given maturity T, the performance measure $\hat{f}^i_{t,T}$ is defined as:

$$\hat{f}^i_{t,T} = -d^i_{t,T}. \tag{9}$$

Thus, the null hypothesis is that no model outperforms the random walk model, i.e. $H_0 : \max_{i=1,\ldots,k} E\left(f^i_T\right) \leq 0$. The test statistic for the observed sample is:

$$\bar{V} = \max_{i=1,\ldots,k}\left\{\sqrt{n}\left(\bar{f}^i_T\right)\right\}, \tag{10}$$

where $\bar{f}^i_T = \sum_{t=1}^n \hat{f}^i_{t,T}/n$. White (2000) suggests that the null hypothesis can be evaluated by applying the stationary bootstrap of Politis and Romano (1994) to the observed values of $\hat{f}^i_{t,T}$.[7] Specifically, B bootstrapped samples of $\hat{f}^i_{t,T}$ are generated. For each bootstrap sample, the following statistic is calculated:

$$\bar{V}_j = \max_{i=1,\ldots,k}\left\{\sqrt{n}\left(\bar{f}^{i*}_{T,j} - \bar{f}_{i,T}\right)\right\}, \tag{11}$$

where $j = 1, 2, \ldots, B$, and $\bar{f}^{i*}_{T,j}$ are the bootstrapped values of $\bar{f}^i_T$. We chose $B = 1000$. White's (2000) reality check p-value is then obtained by comparing $\bar{V}$ with the obtained $\bar{V}_j$ for $j = 1, 2, \ldots, B$ (see White, 2000, p. 1110, for a detailed description).

5.2. Interval forecasts: Statistical testing

Christoffersen's (1998) likelihood ratio test of unconditional coverage is used to evaluate the constructed interval forecasts. The test can be applied for any assumed underlying stochastic process, since is not model dependent (Christoffersen, 1998). The idea of the test is as follows. A sample path $\{F_{t,T}\}_{t=1}^n$ of futures prices is observed for a given maturity, and a series of interval forecasts $\left\{L^i_{t/t-1,T}(1-a), U^i_{t/t-1,T}(1-a)\right\}_{t=1}^T$ is constructed. $L^i_{t/t-1,T}(1-a)$ and $U^i_{t/t-1,T}(1-a)$ denote the lower and upper bounds of a $(1-a)$% interval forecast for time t constructed at time $t-1$ for a given maturity contract based on the ith model, respectively. We test whether the $(1-a)$%-interval forecast is "efficient", i.e., whether the percentage of times that the realized future price at time t falls outside the interval forecast for time t constructed at time $t-1$ is a% for a given maturity. To this end, an indicator function $I^i_{t,T}$ is defined:

$$I^i_{t,T} = \begin{cases} 0, & \text{if } F_{t,T} \in [L^i_{t/t-1,T}(1-a), U^i_{t/t-1,T}(1-a)] \\ 1, & \text{if } F_{t,T} \notin [L^i_{t/t-1,T}(1-a), U^i_{t/t-1,T}(1-a)]. \end{cases} \tag{12}$$

Thus, the null hypothesis of an efficient $(1-a)$% interval forecast $H_0 : E(I^i_{t,T}) = \alpha$ is tested against the alternative $H_1 : E(I^i_{t,T}) \neq \alpha$. Under the null hypothesis, Christoffersen's (1998) test statistic is given by a likelihood ratio test (see Christoffersen, 1998). However, the power of this test may be sensitive to the sample size. Hence, Monte Carlo simulated p-values are generated in order to assess the statistical significance of our results. To calculate the Monte Carlo simulated p-values, a sample of n i.i.d. Bernoulli (a) variables is simulated. Next, for the simulated sample, Christoffersen's (1998) test statistic

[6] In the case of the MCP, the H_1 and H_2 hypotheses are stated as H_1: MCP < 50% and H_2: MCP > 50%.

[7] The stationary bootstrap is applicable to weakly dependent stationary time series; in our case, $\hat{f}_{i,t,T}$ has been found to be stationary. It involves re-sampling *blocks* of a random size from the original time series to form a pseudo time series (or a bootstrapped sample). The block size follows a geometric distribution with a mean block length $1/q$. The main feature of this procedure is that the re-sampled pseudo time series retains the stationarity property of the original series. Following Sullivan et al. (1999), we choose $q = 0.1$, which corresponds to a mean block size of 10. This is a reasonable block size given the low autocorrelation in $\hat{f}_{i,t,T}$ (these results are not reported here). As a robustness check, we have also performed White's (2000) test for alternative mean block sizes of 2, 20, and 30. We found that the results are not sensitive to the choice of the average block size (results not reported here).

is obtained. This procedure is repeated $K = 9999$ times and the empirical distribution of Christoffersen's (1998) test statistic under the null is obtained. Let M be the number of times that the observed test statistic is more extreme than the simulated ones. Then, the Monte-Carlo p-value equals $(1 + M)/(1 + K)$. We construct 99% and 95% interval forecasts to assess the robustness of the obtained results across different levels of significance.

5.3. Point and interval forecasts: Results

In the case of point forecasts, Table 4 shows the RMSE, MAE and MCP values obtained for point forecasts based on the random walk model (Panel A), the economic variables model (Panel B), the AR(2) model (Panel C), the VAR model (Panel D) and the ARMA(1,1) model (Panel E). The results for the equally weighted and weighted combinations of point forecasts are also reported (Panels F and G, respectively). * and ** ($^{+}$ and $^{++}$) denote rejection of the null hypothesis in favor of the alternative H_1 (H_2) at significance levels of 1% and 5%, respectively, by the MDM and ratio tests. We can see that there are 6 combinations of futures series and predictability metrics (out of a total of 54 possible combinations) in which the random walk beats one of the models (i.e., 11% of the cases). On the other hand, the model under consideration outperforms the random walk in 4 out of 54 cases (i.e., 7%). All of these occur under the MCP measure and for the shortest series. Note that under the assumption of the independence of accept/reject decisions, one would expect the models to beat the random walk in only roughly 3 out of 54 cases (i.e., 5% of the cases) at the 5% significance level. Thus, there is weak evidence of a statistically predictable pattern in the evolution of the shortest futures series.

In the case of White's (2000) test, the reality check p-values for the RMSE (MAE) are 0.998 (0.530), 0.999 (0.789) and 0.815 (0.377) for the shortest, second shortest and third shortest series, respectively. Thus, we accept the null hypothesis in all cases. This implies that even the best performing model specification under the RMSE (MAE) metric does not outperform the random walk.

Regarding interval forecasts, Table 5 shows the percentage of observations that fall outside the constructed 99% and 95% interval forecasts, together with Christoffersen's (1998) test statistic values

Table 4
Out-of-sample performance of the model specifications for each of the VIX futures prices series. The root mean squared prediction error (RMSE), mean absolute prediction error (MAE), and mean correct prediction (MCP) of the direction of change in the value of each VIX futures price series are reported. The random walk model (Panel A), the economic variables model (Panel B), the AR(2) model (Panel C), the VAR model (Panel D), and the ARMA(1, 1) model (Panel E) have been implemented. Results for the equally weighted and weighted combination point forecasts are also reported (Panels F and G, respectively). The Modified Diebold-Mariano test (for RMSE and MAE) based on a Newey-West estimator of the variance of $\bar{d}^i_T$ and the ratio test (for MCP) are employed to test the null hypothesis that the random walk and the model under consideration perform equally well. Two alternative hypotheses H_1 and H_2 are considered. H_1: the random walk outperforms the model; and H_2: the model outperforms the random walk. The models have been estimated recursively for the period March 18, 2005, to March 13, 2008.

	Shortest	2nd shortest	3rd shortest
Panel A: Random walk			
RMSE	7.01	5.16	4.55
MAE	4.32	3.31	2.89
Panel B: Economic variables model			
RMSE	7.18	5.25*	4.70
MAE	**4.44****	**3.42****	**3.03****
MCP (%)	**54.67**$^{+}$	47.35	50.94
Panel C: AR(2) model			
RMSE	7.26	5.37	4.81
MAE	4.45	3.46	3.09
MCP (%)	**54.78**$^{+}$	51.66	52.81
Panel D: VAR(1) model			
RMSE	7.60	5.49	4.77
MAE	4.74	3.56	3.06
MCP (%)	52.92	49.74	53.36
Panel E: ARMA(1,1) model			
RMSE	**7.17****	5.23	4.68
MAE	4.34	3.35	2.98
MCP (%)	**55.44**$^{+}$	50.55	52.43
Panel F: Equally weighted combination forecast			
RMSE	7.71	5.55	4.92
MAE	4.82	3.63	3.18
MCP (%)	**53.55**$^{++}$	49.07	51.37

(*continued on next page*)

Table 4 (*continued*)

	Shortest	2nd shortest	3rd shortest
Panel G: Weighted combination forecast			
RMSE	7.77	**5.60****	4.96
MAE	4.90	**3.67****	3.21
MCP (%)	50.09	50.19	50.00

* ($^+$) denotes the rejection of the null hypothesis in favour of the alternative H_1 (H_2) at the 1% significance level.
** ($^{++}$) denotes the rejection of the null hypothesis in favour of the alternative H_1 (H_2) at the 5% significance level.

obtained by the economic variables model, the AR(2) model, the VAR model, the ARMA(1,1) model, the equally weighted combination interval forecasts and the weighted combination interval forecasts (Panels A, B, C, D, E and F, respectively); results are reported for each of the three futures series. * and ** denote the rejection of the null hypothesis at the 1% and 5% significance levels, respectively. We can see that the null hypothesis of efficient interval forecasts is rejected in all instances. This holds for both the 99% and 95% interval forecasts.

6. Out-of-sample evidence: Economic significance

The previously reported results on point forecasts suggest that there is weak evidence of a statistically

Table 5
Statistical efficiency of the bootstrapped interval forecasts. The entries report the percentage of observations that fall outside the bootstrapped intervals, and the values of Christoffersen's (1998) likelihood ratio test of unconditional coverage (LR_{unc}) for each VIX futures price series. The null hypothesis is that the percentage of times that the actually realized futures price falls outside the constructed $(1-\alpha)$% interval forecasts is a%. The results are reported for daily 99% and 95% interval forecasts generated by the economic variables model (Panel A), the AR(2) model (Panel B), the VAR model (Panel C), and the ARMA(1,1) model (Panel D). Results for the equally weighted and weighted combination 99% and 95% interval forecasts are also presented (Panels E and F, respectively). The models have been estimated recursively for the period March 18, 2005, to March 13, 2008.

	Shortest		2nd shortest		3rd shortest	
Interval forecasts	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**
Panel A: Economic variables model interval forecasts						
# Violations (%)	3.16	9.79	2.41	9.15	3.09	8.75
LR_{unc}	**19.96***	**25.37***	**8.95***	**18.33***	**16.50***	**14.23***
Panel B: AR(2) model interval forecasts						
# Violations (%)	3.24	9.88	2.98	8.94**	3.63	9.26
LR_{unc}	**20.67***	**25.58***	**15.63***	**16.16***	**22.98***	**16.98***
Panel C: VAR(1) model interval forecasts						
# Violations (%)	2.92	10.14	4.01	10.47	3.89	9.19
LR_{unc}	**14.30***	**25.25***	**29.92***	**27.84***	**27.54***	**16.93***
Panel D: ARMA(1,1) model interval forecasts						
# Violations (%)	3.38	10.00	2.50	9.86	3.18	9.05
LR_{unc}	**24.05***	**28.09***	**10.30***	**25.07***	**18.22***	**16.77***
Panel E: Equally weighted combination interval forecasts						
# Violations (%)	2.91	10.75	2.61	10.45	4.10	9.77
LR_{unc}	**13.41***	**29.14***	**9.74***	**25.85***	**28.02***	**19.39***
Panel F: Weighted combination interval forecasts						
# Violations (%)	2.91	9.65	3.17	10.26	3.32	9.18
LR_{unc}	**13.41***	**19.91***	**16.22***	**24.27***	**17.32***	**15.27***

* Denotes rejection of the null hypothesis at the 1% significance level.
** Denotes rejection of the null hypothesis at the 5% significance level.

predictable pattern in the evolution of the shortest futures series based on the MDM test. Moreover, none of the bootstrapped 99% and 95% interval forecasts were found to be efficient. To provide a definite answer on the issue of predictability in volatility futures markets, the economic significance of the forecasts obtained is assessed by applying trading strategies based on point and interval forecasts. The trading strategies are followed despite the fact that there is no evidence of a statistically predictable pattern. This is because the statistical evidence does not always corroborate that of a financial criterion (see, e.g., Ferson, Sarkissian, & Simin, 2003). The trading strategies involve a single volatility futures contract. Transaction costs have been taken into account; the standard transaction fee in the VIX futures market is \$0.50 per transaction, which represents 0.003% of the contract value (on average) for each futures series under consideration.

6.1. Testing for economic significance: Measures of performance

The profitability of the trading strategies is evaluated in terms of the Sharpe Ratio (SR) and Leland's (1999) alpha (A_p). The statistical significance of the two performance measures is assessed by bootstrapping their 95% confidence intervals. To this end, the stationary bootstrap of Politis and Romano (1994) has been employed.[8] The continuously compounded one month Libor rate is used as the risk free rate to calculate both measures of performance.

Leland's (1999) alpha is employed in order to account for the presence of non-normality in the distribution of the trading strategy's returns. It is defined as:

$$A_p = E\left(r_p\right) - B_p\left[E\left(r_{mkt}\right) - r_f\right] - r_f, \tag{13}$$

where r_p is the return on the trading strategy, r_f is the risk-free rate of interest, r_{mkt} is the return on the market portfolio, $B_p = \frac{\mathrm{cov}\left(r_p, -(1+r_{mkt})^{-\gamma}\right)}{\mathrm{cov}\left(r_{mkt}, -(1+r_{mkt})^{-\gamma}\right)}$ is a measure of risk similar to the CAPM's beta, and $\gamma = \frac{\ln[E(1+r_{mkt})]-\ln\left(1+r_f\right)}{\mathrm{var}[\ln(1+r_{mkt})]}$ is a measure of risk aversion.

A two step procedure is employed for calculating Leland's alpha. First, γ and B_{str} are computed for each time step. We use the one month continuously compounded Libor rate and the return on the S&P 500 as proxies for r_f and r_{mkt}, respectively. Second, the following regression is estimated:

$$r^i_{p,t} - B^i_{p,t}\left[r_{mkt,t} - r_{f,t}\right] - r_{f,t} = A^i_p + \varepsilon_t, \tag{14}$$

where $r^i_{p,t}$ and A^i_p are the return on the trading strategy and Leland's alpha, respectively, which are based on the forecasts from the ith model ($i = 1$ (economic variables model), 2 (AR(2) model), 3 (VAR model), 4 (ARMA(1,1) model), 5 (equally weighted combination forecast), 6 (weighted combination forecast)). If $A^i_p > 0$ then we conclude that the trading strategy offers an expected return in excess of its equilibrium risk adjusted level.

6.2. Trading strategy and results based on point forecasts

The economic significance of the point forecasts constructed is evaluated in terms of the following trading rule:

if $F_{t-1,T} < (>) \hat{F}^i_{t|t-1,T}$, then go long (short);

if $F_{t-1,T} = \hat{F}^i_{t|t-1,T}$, then do nothing.

The rationale of this trading rule is as follows: if the current futures price is higher (lower) than the forecasted futures price, then the price is anticipated to decrease (increase) and the investor goes short (long). If the current futures price is equal to the forecasted futures price, then the investor takes no action and maintains his/her position.

Table 6 shows the annualised SR and A_p, together with their respective bootstrapped 95% confidence intervals (95% CI), for the three VIX futures series. Results are reported for the trading strategy based on point forecasts derived by the economic variables model (Panel A), the AR(2) model (Panel B), the VAR model (Panel C) and the ARMA(1, 1) model

[8] We use the stationary bootstrap method to obtain the confidence intervals for the SR and the alpha estimates so as to take into account the non-normality of the returns of the trading strategies; these exhibit excess kurtosis and skewness that range from 6 to 13 and from −1 to 2 respectively, across the three futures maturities. The non-normality of volatility futures returns is consistent with previous findings in the related literature for other futures markets (see e.g. Taylor, 1985, and references therein). Given the untabulated low autocorrelation in excess returns, the average block size was chosen to be 10 (i.e. $q = 0.1$). As a robustness check, we have also constructed bootstrapped confidence intervals for alternative mean block sizes of 2, 20, and 30. We found that the results on SR and alpha are robust to the choice of q.

Table 6
Trading strategy with VIX futures based on point forecasts from March 18, 2005, to March 13, 2008. The entries show the annualised Sharpe ratio (SR) and Leland's (1999) alpha (A_p), together with their respective bootstrapped 95% confidence intervals (95% CI) in parentheses. The stationary bootstrap of Politis and Romano (1994) has been employed. The strategy is based on point forecasts obtained from the economic variables model (Panel A), the AR(2) model (Panel B), the VAR model (Panel C) and the ARMA(1, 1) model (Panel D). Results for the equally weighted and weighted combination point forecasts are also reported (Panels E and F, respectively). The SR for a naïve buy and hold strategy in VIX volatility futures is 0.0178 (95% CI = (−0.07, 0.08)) for the shortest maturity series, 0.0421 (95% CI = (−0.02, 0.10)) for the second shortest maturity series, and 0.0532 (95% CI = −0.01, 0.12)) for the third shortest maturity series.

	Shortest	2nd shortest	3rd shortest
Panel A: Economic variables model point forecasts			
Sharpe ratio	0.085	0.013	−0.015
95% CI	*(0.02, 0.14)*	*(−0.07, 0.09)*	*(−0.09, 0.06)*
Leland's Alpha	0.854	0.104	−0.101
95% CI	*(0.21, 1.51)*	*(−0.46, 0.69)*	*(−0.59, 0.40)*
Panel B: AR(2) model point forecasts			
Sharpe ratio	0.017	−0.013	−0.031
95% CI	*(−0.05, 0.09)*	*(−0.08, 0.06)*	*(−0.11, 0.05)*
Leland's alpha	0.167	−0.071	−0.181
95% CI	*(−0.58, 0.92)*	*(−0.57, 0.49)*	*(−0.63, 0.32)*
Panel C: VAR(1) model point forecasts			
Sharpe ratio	−0.018	−0.039	0.048
95% CI	*(−0.11, 0.07)*	*(−0.11, 0.04)*	*(−0.03, 0.12)*
Leland's alpha	−0.170	−0.270	0.300
95% CI	*(−1.1, 0.68)*	*(−0.88, 0.27)*	*(−0.18, 0.78)*
Panel D: ARIMA(1, 1, 1) model point forecasts			
Sharpe ratio	0.013	−0.007	−0.016
95% CI	*(−0.07, 0.10)*	*(−0.09, 0.07)*	*(−0.09, 0.06)*
Leland's alpha	0.138	−0.034	−0.095
95% CI	*(−0.69, 0.89)*	*(−0.68, 0.53)*	*(−0.47,0.39)*
Panel E: Equally weighted point forecasts			
Sharpe ratio	0.020	−0.019	−0.033
95% CI	*(−0.06, 0.11)*	*(−0.10, 0.06)*	*(−0.11, 0.04)*
Leland's alpha	0.196	−0.120	−0.195
95% CI	*(−0.53, 0.91)*	*(−0.67, 0.52)*	*(−0.66, 0.31)*
Panel F: Weighted point forecasts			
Sharpe ratio	−0.011	−0.050	−0.073
95% CI	*(−0.08, 0.06)*	*(−0.13, 0.02)*	*(−0.13, −0.01)*
Leland's alpha	−0.104	−0.343	−0.447
95% CI	*(−0.81, 0.57)*	*(−0.93, 0.26)*	*(−0.83, −0.04)*

(Panel D), as well as the equally weighted (Panel E) and weighted (Panel F) combination point forecasts. We can see that SR and A_p are insignificant in all cases but one. This implies that almost all trading strategies based on point forecasts do not yield economically significant profits. The results are similar to those obtained for a naïve buy and hold strategy in VIX futures, which yields a SR equal to 0.0178 (95% CI = $(-0.07, 0.08)$) for the shortest series, 0.0421 (95% CI = $(-0.02, 0.10)$) for the second shortest series, and 0.0532 (95% CI = $-0.01, 0.12)$) for the third shortest series.

6.3. Trading strategy and results based on interval forecasts

The economic significance of the bootstrapped interval forecasts is evaluated in terms of the following trading rule:

$$\text{if } F_{t-1,T} < (>) \frac{U^i_{t|t-1,T}(1-\alpha) + L^i_{t|t-1,T}(1-\alpha)}{2},$$

then go long (short);

$$\text{if } F_{t-1,T} = \frac{U^i_{t|t-1,T}(1-\alpha) + L^i_{t|t-1,T}(1-\alpha)}{2},$$

then do nothing.

The rationale behind this trading rule is as follows: if the futures price is closer to the lower (upper) bound of the next day's interval forecasts, then we anticipate that the index price will increase (decrease), and as a result the investor should go long (short).

Table 7 shows the annualised SR and A_p, together with their respective bootstrapped 95% confidence intervals (95% CI), for the three VIX futures series. Results are reported for the trading strategy based on 99% and 95% bootstrapped interval forecasts derived by the economic variables model (Panel A), the AR(2) model (Panel B), the VAR model (Panel C) and the ARMA model (Panel D), as well as the weighted (Panel E) and equally weighted (Panel F) combination interval forecasts. We can see that the results are similar for the strategies based on the 99% and 95% bootstrapped interval forecasts. In particular, the SR and A_p are insignificant in all cases. This means that, overall, the trading strategies based on bootstrapped interval forecasts do not yield significant profits, just as was the case with the trading strategies based on point forecasts. The results are similar to those obtained for a naïve buy and hold strategy in VIX volatility futures. In particular, the SR equals 0.0178 [95% CI $= (-0.07, 0.08)$] for the shortest series, 0.0421 [95% CI $= (-0.02, 0.10)$] for the second shortest series, and 0.0532 [95% CI $= -0.01, 0.12)$] for the third shortest series.

7. Conclusions

This paper has investigated, for the first time, whether volatility futures prices per se can be forecasted. To this end, the most liquid volatility futures market (futures on VIX) has been considered. A number of alternative model specifications have been employed: the economic variables model, the AR(2) model, the VAR model and the ARMA(1,1) model. Equally weighted and weighted combination forecasts have also been considered. Both point and bootstrapped interval forecasts have been constructed and their statistical and economic significance have been evaluated. The latter is assessed by means of trading strategies using the VIX futures. This has implications for the efficiency of the VIX volatility futures market.

Regarding the statistical significance of the forecasts obtained, in the case of point forecasts, we found weak evidence of a statistically predictable pattern in the evolution of the shortest futures series. In the case of the interval forecasts, no model specification had predictive power. Regarding the economic significance of the forecasts obtained, the constructed forecasts did not yield economically significant profits.

Overall, our results imply that one cannot reject the hypothesis that the VIX volatility futures market is informationally efficient. These findings are consistent with those of Konstantinidi et al. (2008), who studied the efficiency of the VIX futures market indirectly. On the other hand, our results are in contrast to those found about the efficiency of other futures markets (stock, currency, interest rate and commodities), where predictability in either statistical or economic terms has been documented. However, the fact that the VIX futures market is found to be efficient does not invalidate the trading of VIX futures. This is because VIX futures can also be used for hedging against changes in volatility. After all, this was the main motivation for their introduction (see Brenner & Galai, 1989, 1993).

Future research should investigate the issue of predictability in volatility futures markets at longer horizons. It has been well documented that the predictability in asset returns increases as the horizon increases (see, e.g. Poterba & Summers, 1988). However, a longer horizon study is beyond the scope of this paper due to data limitations, as the VIX market has only been operating since 2004. Intra-day data should also be used to test whether any predictable patterns can be detected within the data; this will be particularly useful for day-traders. Finally, it may be worth considering

Table 7
Trading strategy with VIX futures based on bootstrapped interval forecasts from March 18, 2005, to March 13, 2008. The entries show the annualised Sharpe ratio (SR) and Leland's (1999) alpha (A_p), with their respective bootstrapped 95% confidence intervals (95% *CI*) in parentheses. The stationary bootstrap of Politis and Romano (1994) has been employed. The strategy is based on 99% and 95% bootstrapped interval forecasts obtained from the economic variables model (Panel A), the AR(2) model (Panel B), the VAR model (Panel C) and the ARMA(1,1) model (Panel D). Results for the equally weighted and weighted combination point forecasts (Panels E and F, respectively) are also reported. The SR for a naïve buy and hold strategy in VIX volatility futures is 0.0178 (95% CI = (−0.07, 0.08)) for the shortest maturity series, 0.0421 (95% CI = (−0.02, 0.10)) for the second shortest maturity series, and 0.0532 (95% CI = (−0.01, 0.12)) for the third shortest maturity series.

	Shortest		2nd shortest		3rd shortest	
Interval forecasts	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**
Panel A: Economic variable model interval forecasts						
Sharpe ratio	0.020	−0.001	0.036	0.034	0.007	0.029
95% CI	*(−0.05, 0.08)*	*(−0.08, 0.06)*	*(−0.02, 0.10)*	*(−0.03, 0.10)*	*(−0.07, 0.08)*	*(−0.03, 0.09)*
Leland's alpha	0.194	−0.014	0.267	0.244	0.050	0.181
95% CI	*(−0.50, 0.91)*	*(−0.67, 0.64)*	*(−0.20, 0.73)*	*(−0.24, 0.73)*	*(−0.43, 0.54)*	*(−0.21, 0.59)*
Panel B: AR(2) model interval forecasts						
Sharpe ratio	0.018	0.016	0.018	0.044	0.063	0.035
95% CI	*(−0.04, 0.08)*	*(−0.05, 0.07)*	*(−0.05, 0.08)*	*(−0.02, 0.11)*	*(−0.01, 0.13)*	*(−0.03, 0.10)*
Leland's alpha	0.178	0.163	0.131	0.321	0.395	0.224
95% CI	*(−0.44, 0.82)*	*(−0.42, 0.76)*	*(−0.35, 0.61)*	*(−0.16, 0.80)*	*(−0.08, 0.84)*	*(−0.23, 0.67)*
Panel C: VAR(1) model interval forecasts						
Sharpe ratio	0.035	0.014	0.026	0.010	0.034	0.026
95% CI	*(−0.03, 0.09)*	*(−0.05, 0.07)*	*(−0.04, 0.09)*	*(−0.06, 0.07)*	*(−0.05, 0.11)*	*(−0.04, 0.10)*
Leland's alpha	0.346	0.142	0.183	0.080	0.210	0.167
95% CI	*(−0.23, 0.96)*	*(−0.41, 0.70)*	*(−0.32, 0.63)*	*(−0.39, 0.55)*	*(−0.19, 0.67)*	*(−0.25, 0.65)*
Panel D: ARMA(1, 1, 1) model interval forecasts						
Sharpe ratio	0.025	0.022	0.054	0.034	0.038	0.032
95% CI	*(−0.04, 0.08)*	*(−0.04, 0.09)*	*(−0.02, 0.12)*	*(−0.03, 0.10)*	*(−0.03, 0.11)*	*(−0.04, 0.10)*
Leland's alpha	0.246	0.220	0.390	0.249	0.236	0.196
95% CI	*(−0.34, 0.93)*	*(−0.38, 0.87)*	*(−0.14, 0.84)*	*(−0.26, 0.75)*	*(−0.18, 0.68)*	*(−0.22, 0.64)*
Panel E: Equally weighted combination interval forecasts						
Sharpe ratio	0.029	0.019	0.033	0.034	0.036	0.025
95% CI	*(−0.03, 0.09)*	*(−0.04, 0.07)*	*(−0.03, 0.09)*	*(−0.03, 0.10)*	*(−0.03, 0.10)*	*(−0.05, 0.10)*
Leland's alpha	0.288	0.191	0.240	0.251	0.223	0.162
95% CI	*(−0.33, 0.91)*	*(−0.35, 0.77)*	*(−0.25, 0.70)*	*(−0.23, 0.75)*	*(−0.16, 0.62)*	*(−0.31, 0.62)*
Panel F: Weighted combination interval forecasts						
Sharpe ratio	−0.016	0.017	0.061	0.031	0.013	0.031
95% CI	*(−0.08, 0.04)*	*(−0.05, 0.08)*	*(−0.004, 0.12)*	*(−0.03, 0.09)*	*(−0.06, 0.08)*	*(−0.04, 0.10)*
Leland's alpha	−0.159	0.171	0.439	0.222	0.085	0.191
95% CI	*(−0.79, 0.44)*	*(−0.50, 0.82)*	*(−0.03, 0.95)*	*(−0.25, 0.66)*	*(−0.36, 0.53)*	*(−0.24, 0.63)*

more complex model specifications, given that the answer on the predictability question always depends on the assumed specification of the predictive regression.

Acknowledgements

We are grateful to Graham Elliott (the co-editor), one Associate Editor and two anonymous referees for their stimulating, thorough and constructive comments. We would like also to thank Mascia Bedendo, Ales Cerny, Alexandros Kostakis, Dimitris Malliaropoulos, Bruce McCullough, Joëlle Miffre, Marcus Nossman, Dimitris Politis, Allan Timmermann, Halbert White, and participants at the 2009 Financial Management Association European Conference (Turin) for useful discussions and comments. Any remaining errors are our responsibility alone.

References

Bakshi, G., Cao, C., & Chen, Z. (2000). Do call prices and the underlying stock always move in the same direction? *Review of Financial Studies*, *13*, 549–584.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*, 451–468.

Bessembinder, H., & Chan, K. (1992). Time varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics*, *32*, 169–193.

Brenner, M., & Galai, D. (1989). New financial instruments for hedging changes in volatility. *Financial Analysts Journal*, *45*, 61–65.

Brenner, M., & Galai, D. (1993). Hedging volatility in foreign currencies. *Journal of Derivatives*, *1*, 53–59.

Brenner, M., Shu, J., & Zhang, J. E. (2008). *The market for volatility trading; VIX futures*. Working Paper, New York University, Stern School of Business.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*, 841–862.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.

Coppola, A. (2008). Forecasting oil price movements: Exploiting the information in the futures market. *Journal of Futures Markets*, *28*, 34–56.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*, 1057–1072.

Diebold, F. X., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–265.

Dotsis, G., Psychoyios, D., & Skiadopoulos, G. (2007). An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking and Finance*, *31*, 3584–3603.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, *46*, 555–576.

Fama, E. F., & French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums and the theory of storage. *Journal of Business*, *60*, 55–73.

Ferson, W. E., Sarkissian, S., & Simin, T. T. (2003). Spurious regressions in financial economics. *Journal of Finance*, *58*, 1393–1413.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*, 197–204.

Grünbichler, A., & Longstaff, F. A. (1996). Valuing futures and options on volatility. *Journal of Banking and Finance*, *20*, 985–1001.

Hartzmark, M. L. (1987). Returns to individual traders of futures: Aggregate results. *Journal of Political Economy*, *95*, 1292–1306.

Hartzmark, M. L. (1991). Luck versus forecast ability: Determinants of trader performance in futures markets. *Journal of Business*, *64*, 49–74.

Harvey, D. I., Leybourne, S. J., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Jarrow, R., Li, H., & Zhao, F. (2007). Interest rate caps smile too! But can the LIBOR market models capture it? *Journal of Finance*, *62*, 345–382.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, *6*, 95–101.

Jiang, G., & Oomen, R. (2001). *Hedging derivatives risks: A simulation study*. Working Paper, University of Arizona.

Kearns, J., & Manners, P. (2004). *The profitability of speculators in currency futures markets*. Working Paper, Reserve Bank of Australia.

Kellard, N., Newbold, P., Rayner, T., & Ennew, C. (1999). The relative efficiency of commodity futures markets. *Journal of Futures Markets*, *19*, 413–432.

Kho, B. C. (1996). Time-varying risk premia, volatility, and technical trading rule profits: Evidence from foreign currency futures markets. *Journal of Financial Economics*, *41*, 249–290.

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

Leland, H. E. (1999). Beyond mean-variance: Performance measurement in a nonsymmetrical world. *Financial Analysts Journal*, *55*, 27–35.

Lin, Y. (2007). Pricing VIX futures: Evidence from integrated physical and risk-neutral probability measures. *Journal of Futures Markets*, *27*, 1175–1217.

Miffre, J. (2001a). Efficiency in the pricing of the FTSE 100 futures contracts. *European Financial Management*, *7*, 9–22.

Miffre, J. (2001b). Economic activity and time variation in expected futures returns. *Economic Letters*, *73*, 73–79.

Miffre, J. (2002). Economic significance of the predictable movements in futures returns. *Economic Notes by Banca Monte dei Paschi di Siena SpA*, *31*, 125–142.

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

Nossman, M., & Vilhelmsson, A. (2009). Is the VIX futures market able to predict the VIX index? A test of the expectations hypothesis. *Journal of Alternative Investments*, *12*, 54–67.

Pascual, L., Romo, J., & Ruiz, E. (2001). Effects of parameter estimation on prediction densities: A bootstrap approach. *International Journal of Forecasting*, *17*, 83–103.

Politis, D., & Romano, J. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*, 1303–1313.

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

Poterba, J., & Summers, L. (1988). Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics*, *22*, 27–60.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance*, *44*, 1115–1153.

Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, *54*, 1647–1691.

Taylor, S. J. (1985). The behavior of futures prices over time. *Applied Economics*, *17*, 713–734.

Taylor, S. J. (1992). Rewards available to currency futures speculators: Compensation for risk or evidence of inefficient pricing? *Economic Record*, *68*(Supplement), 105–116.

Wang, C. (2004). Futures trading activity and predictable foreign exchange market movements. *Journal of Banking and Finance*, *28*, 1023–1041.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*, 1455–1508.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*, 1097–1126.

Yoo, J., & Maddala, G. S. (1991). Risk premia and price volatility in futures markets. *Journal of Futures Markets*, *11*, 165–177.

Zhang, J. E., & Zhu, Y. (2006). VIX futures. *Journal of Futures Markets*, *26*, 521–531.

Eirini Konstantinidi is a final year doctoral student at the University of Piraeus in the Department of Banking and Financial Management. She holds a B.Sc. in Statistics from the Athens University of Economics and Business (AUEB), an M.Sc. in International Accounting and Financial Studies from the University of Strathclyde and an M.B.A. from AUEB. A paper from her doctoral thesis has been published in the *Journal of Banking and Finance*. She has presented her work in doctoral seminars and international conferences, and has acted as a referee for academic journals. She has been awarded a doctoral student travel grant by the American Finance Association to attend the American Finance Association's 2009 Annual Meeting.

George Skiadopoulos is Associate Professor (elected) in the Department of Banking and Financial Management of the University of Piraeus. He is also an Associate Research Fellow at the Financial Options Research Centre of the University of Warwick. He has published in various academic journals and books, such as *Energy Economics*, *European Financial Management Journal*, *International Journal of Theoretical and Applied Finance, Journal of Alternative Investments, Journal of Banking and Finance, Journal of Futures Markets, Journal of Risk Finance, RISK, Review of Derivatives Research*, etc., and is a frequent speaker in international conferences. He is a member of the editorial board of the *Journal of Business Finance and Accounting* and serves in the Academic Advisory Council of the Professional Risk Managers International Association (PRMIA).