International Journal of Forecasting 32 (2016) 527–547

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

On the predictability of model-free implied correlation

Chryssa Markopoulou[a,*], Vasiliki Skintzi[b], Apostolos Refenes[a]

[a] Athens University of Economics and Business, Athens, Greece
[b] University of Peloponnese, Tripolis, Greece

ARTICLE INFO

Keywords:
Predictability
Implied correlation
Combining forecasts
Market efficiency
Option strategy

ABSTRACT

This paper investigates the existence of predictable patterns in the evolution of the implied correlation series. To this end, alternative time-series specifications are employed to model the correlation dynamics, and the statistical and economic significance of out-of sample forecasts is assessed. The statistical measures provide strong evidence in favor of predictable patterns in the S&P 100 options market. A trading strategy designed to exploit daily changes in the series can yield abnormal profits; however, these profits disappear when transaction costs are incorporated. We conclude that the efficient market hypothesis cannot be rejected.

0. Introduction

The existing literature has focused primarily on the ability of risk-neutral moments extracted from option prices to provide accurate forecasts of the corresponding realized moments, or to improve asset allocation and pricing decisions. Another important yet distinct question is whether risk-neutral moments are predictable per se. The existence of predictable patterns in the dynamics of risk-neutral moments raises questions regarding the efficiency of the option markets, and enables market participants to develop profitable trading strategies. Thus far, only a limited number of studies have dealt with the predictability of option-implied measures, primarily volatility. The current paper fills this research gap by undertaking a comprehensive study of the dynamics and the unexplored predictable patterns in the evolution of the implied correlation series.[1]

It is important for both academics and practitioners to understand the dynamics of risk-neutral correlations. The implied correlation is a forward-looking measure of the aggregate stock market diversification that, according to Skintzi and Refenes (2005), also quantifies the difference between the minimum and maximum values of the portfolio variance. Moreover, a risk-neutral measure of the correlation may provide important information in an asset-pricing and asset-allocation framework. To this end, Cosemans (2011) and Driessen, Maenhout, and Vilkov (2012) provide evidence that risk-neutral correlation measures can explain the changes in expected returns over time. From a more practical perspective, market participants can exploit forecasts of the risk-neutral correlation in order to form profitable trading strategies. Volatility and correlation trading strategies have attracted the interest of investors, particularly given the dramatic

* Corresponding author.
E-mail address: cmarkopoulou@aueb.gr (C. Markopoulou).

1 Existing studies that focus on implied correlations examine a different question, namely the accuracy of implied correlation measures in forecasting expected correlations for various asset classes (e.g., Campa & Chang, 1998, and Lopez & Walter, 2000, for exchange rates; Han, 2007, for interest rates; and Skintzi & Refenes, 2005, for equities). Moreover, DeMiguel, Plyakha, Uppal, and Vilkov (2013) examine whether option-implied correlations can improve asset allocation decisions, whereas Buss and Vilkov (2012) and Chang, Christoffersen, Jacobs, and Vainberg (2012) use option-implied information to derive equity betas.

http://dx.doi.org/10.1016/j.ijforecast.2015.09.008
0169-2070/

increase in stock market volatilities and correlations that followed the 2008 financial crisis. Interestingly, in July 2009, the Chicago Board Options Exchange (CBOE) launched the S&P 500 implied correlation index for measuring the expected average correlation between the price returns of the index components implied by S&P 500 index option prices and the prices of options on the 50 largest components of S&P 500.[2]

In the context of volatility, David and Veronesi (2002) and Guidolin and Timmermann (2003) provide a theoretical explanation as to why the implied volatility may by predictable by linking option prices and the implied volatility to economic uncertainty. In the context of correlation, Buraschi, Trojani, and Vedolin (2013) develop a structural equilibrium model that links the differential pricing of index and individual equity options to aggregate economic uncertainty and the diversity in beliefs across investors.

A number of studies have investigated whether the dynamics of the implied moments of distribution are predictable. Goyal and Saretto (2009), Harvey and Whaley (1992), and Konstantinidi, Skiadopoulos, and Tzagkaraki (2008) all suggest that there is significant predictability in the dynamics of the implied volatility; nevertheless, the economic significance of these results is mixed. Bernales and Guidolin (2014) and Goncalves and Guidolin (2006) examine whether the implied volatility surface contains any patterns that can be exploited. Although these researchers discern statistically significant predictable patterns, the trading profits yielded by these patterns are not significant once transaction costs are taken into account. Neumann and Skiadopoulos (2013) exploit predictable patterns in the dynamics of option-implied higher-order moments. To the best of our knowledge, the only related study that explores the dynamics of the risk-neutral correlation is that conducted by Härdle and Silyakova (2012). However, their study uses only one model, a dynamic semi-parametric factor model, to capture the dynamics of the implied correlation surface and to forecast a future implied correlation in the German market.

Our study extends the literature on the predictability of option-implied moments and makes several contributions to the predictability of the risk-neutral correlation, and thus, to the informational efficiency of the S&P 100 options market. First, we examine whether the dynamics of the implied correlation series, as a measure of the market-wide correlation, contain any predictable pattern. Toward this end, we fit alternative time series specifications for modeling and forecasting the dynamics of the series. We obtain out-of-sample forecasts and assess their performances based on statistical evaluation criteria. Second, we investigate whether market participants can gain significant abnormal returns by exploiting the predictability of the implied correlation series. Based on the notion of dispersion trade, we build a trading strategy that is exposed to correlation risk and that trades inverse positions in index options and stock component options. Third, an extensive dataset consisting of the S&P 100 index options and the individual options of the stocks that constitute the S&P 100 allows us to evaluate the dynamics of the implied correlation process during several periods of financial turmoil. The sample period extends from January 1996 to October 2010, a period that encompasses several turbulent periods associated with high volatilities and low returns, such as the increased market volatility after the attacks on September 11, 2001, and the US subprime mortgage crisis that resulted in the recent global financial turmoil. Finally, we derive the implied correlation measure without relying on an option-pricing model and diverge from similar studies by adjusting the observed American option prices to account for an early exercise premium.

The risk-neutral correlation measure examined in this study uses the notion of 'equicorrelation'. More specifically, the equicorrelation assumption is necessary when, within the context of several financial applications, estimates and forecasts of an entire correlation matrix are required as inputs.[3] Although this assumption may appear to be excessively restrictive for the applicability of the index, it has several important advantages. In an early study of asset allocation, Elton and Gruber (1973) found that the assumption of equicorrelation reduces estimation errors and provides superior portfolio selection. More recently, Pollet and Wilson (2010) provided a theoretical explanation and empirical evidence showing that the average realized stock market correlation predicts future stock market returns. Furthermore, Engle and Kelly (2012) suggest that multivariate models based on the equicorrelation assumption improve the portfolio allocation relative to unrestricted models. Within the context of implied correlations, Skintzi and Refenes (2005) interpret the implied correlation index as the market view of future stock market diversification, and find that implied correlation measures outperform historical ones in predicting future correlations. Moreover, Driessen et al. (2012) indicate that the average implied correlation has significant predictive power for future stock market returns, and construct a correlation trading strategy that aims to exploit the correlation risk premium.

The empirical results of our study, based on alternative time series specifications, suggest that there is a strong, predictable pattern in the evolution of the implied correlation structure, thus casting doubt on the efficiency of the options market. Turning our attention to the economic significance of the forecasts obtained, we find that the trading strategy implemented yields abnormal profits that vanish when transaction costs are considered. We conclude that the informational efficiency of the S&P 100 market cannot be rejected. Our results are robust across different in-sample sizes and forecast periods.

2 The CBOE S&P500 Implied Correlation Index was not employed for the empirics of this paper because it is not a model-free constant maturity index. Moreover, its use in this context would induce liquidity and microstructure-related biases in our results, given that no filtering rules are used on the options involved in its calculation. Comparable previous studies (Driessen et al., 2012) have also constructed an implied correlation index from scratch rather than using an existing metric.

3 The equicorrelation assumption is only relevant when the implied correlation measure is used as a forecast of individual pair-wise correlations, and does not refer to the construction of the index.

The remainder of this paper is organized as follows: Section 1 describes the methodology for constructing the model-free implied correlation (MFIC), the dataset used for calculating the MFIC series, and the descriptive statistics of the series under examination. Section 2 presents the alternative model specifications used for forecasting purposes. Finally, Sections 3 and 4 discuss the in-sample and out-of-sample performances of the models under consideration and the trading strategy, respectively. We then assess the robustness of our results for different sample sizes and forecasting horizons and present the findings in Section 5. We present our summary and conclusions in Section 6.

1. Methodology and data

Intuitively, the variance of an index depends on the variances of the constituent stocks and the pairwise correlations, defined as follows:

$$\sigma_{I,t}^2 = \sum_{i=1}^{N} w_{i,t}^2 \sigma_{i,t}^2 + 2\sum_{i=1}^{N}\sum_{j>i} w_{i,t} w_{j,t} \rho_{ij,t} \sigma_{i,t} \sigma_{j,t}, \tag{1}$$

where $\sigma_{I,t}^2$ is the index variance; $\sigma_{i,t}$ is the standard deviation of asset i, $i = 1, \ldots, N$; $\rho_{ij,t}$ is the correlation between assets i and j; and $w_{i,t}$ is the relative weight of each index component i, with all variables being taken at time t.

The correlation measure examined in this study is a weighted average of all pair-wise correlations of the constituents of the index, defined as

$$\bar{\rho}_t = \rho_{P,t} = \sum_{i=1}^{N}\sum_{j>i} c_{ij,t} \rho_{ij,t},$$
$$\text{where } c_{ij,t} = \frac{w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t}}{\sum_{i=1}^{N}\sum_{j>i} w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t}}. \tag{2}$$

Thus, $\bar{\rho}_t$ is an average measure of the degree of diversification in the market represented by the index (see Skintzi & Refenes, 2005), and can also be written as

$$\bar{\rho}_t = \frac{\sigma_{I,t}^2 - \sum_{i=1}^{N} w_{i,t}^2 \sigma_{i,t}^2}{2\sum_{i=1}^{N-1}\sum_{j\neq i} w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t}}. \tag{3}$$

The implied correlation measure is derived from Eq. (3), where the only variables required are the volatilities of the assets. Historical volatility and correlation estimators rely on the assumption that the future dynamics of the series will be similar to those in the past. To overcome the ambiguities caused by this assumption, index and stock return volatilities are inferred from currently traded option prices. Naturally, option prices reflect the current market view of future price movements of the underlying asset. Therefore, the estimation of asset price distribution moments from option prices is widely considered to outperform historical estimates of the volatility in terms of informational efficiency.

We compute Eq. (3) using model-free implied volatility measures proposed in the recent literature (e.g., Jiang & Tian, 2005). However, the methodology assumes that the underlying options are European-style, whereas our dataset, which is described thoroughly below, consists of American-style options. We address this issue and correct the observed American option prices for the early exercise premium (EEP) using the methodology proposed by Tian (2011). The procedure used to derive the model-free implied volatilities from option prices and to extract the EEP is described thoroughly in Appendix A. The resulting implied correlation is essentially model-free, as the volatility estimates are deduced from the observed prices without relying on any specific option-pricing model. Throughout the paper, we refer to the inferred correlation measure as the model-free implied correlation (MFIC).

For the purposes of our study, daily closing quotes for options on the S&P 100 index and the constituent stocks are used. Our sample, which extends from January 1996 to October 2010, includes 179 stocks. Data on option prices, underlying asset prices and company distributions are retrieved from OptionMetrics. In addition, the zero-coupon interest rate curve, provided by the OptionMetrics database, is used as the risk-free interest rate. Any additions to or deletions from the index are made as needed whenever any of the S&P's inclusion criteria are violated. We retrieve the list of constituent stocks from Datastream. On a daily basis, we compute the actual daily weight of stock i based on its daily market capitalization divided by the total market capitalization of the stocks included in the index on the specific day.[4] The subset from January 4, 1996, to December 31, 2000, will be used for the in-sample estimation, whereas the remaining period will be retained for the out-of-sample evaluation of the forecasting methods.

Several filtering rules are applied. First, to ensure trading activity, we choose options with positive bid prices and open interest. Second, we eliminate options with less than one week to maturity, as they are more vulnerable to liquidity and microstructure issues. Third, we discard options with implied volatilities that are either greater than one or missing.[5] In addition, options with extreme moneyness levels are eliminated; our analysis includes only call options with deltas greater than 0.15 and smaller than 0.5 and put options with deltas greater than −0.5 and less than −0.05. Finally, we exclude options that violate arbitrage bounds.

[4] Our analysis does not suffer from a survivorship bias because it is not limited to options written on stocks that have been traded continuously throughout our sample period. Instead, our MFIC index replicates the composition of the S&P100 index on a daily basis. The main criteria for a company to be included in the S&P 100 index are market capitalization, liquidity and the availability of individual stock options. As a result, heavily traded and liquid options introduced during the sample period are expected to be included in the index.

[5] OptionMetrics does not report the implied volatility for an option in the case of a special settlement, a failure of the implied volatility process to converge, the unavailability of the underlying asset price, a vega value of below 0.5, or the midpoint of the bid/ask prices being below the intrinsic value.

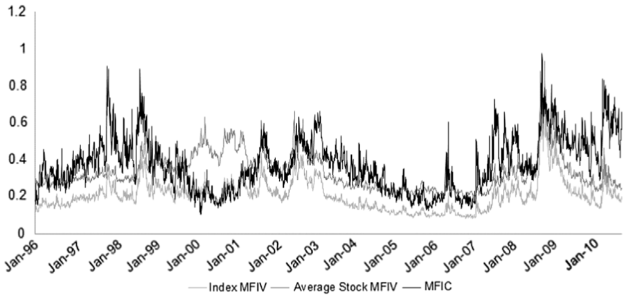

Fig. 1. Model-free implied correlation (MFIC), and model-free implied volatility series for the index and the individual stocks. The time series of the daily model-free implied correlation (MFIC), the time series of the daily cross-sectional weighted average of the model-free implied volatility for the S&P 100 constituent stocks (average stock MFIV), and the time series of the daily MFIV for the S&P 100 index (index MFIV) for the period from January 3, 1996, to October 29, 2010.

Table 1
Summary statistics.

	1996–2000		2001–2010		1996–2010	
	MFIC	Transformed MFIC	MFIC	Transformed MFIC	MFIC	Transformed MFIC
Mean	0.360	−0.617	0.379	−0.537	0.373	−0.564
Median	0.349	−0.622	0.359	−0.578	0.356	−0.594
Maximum	0.907	2.281	0.977	3.756	0.977	3.756
Minimum	0.106	−2.132	0.126	−1.934	0.106	−2.132
Std. Dev.	0.135	0.622	0.147	0.672	0.143	0.656
Skewness	0.808	0.609	0.492	0.396	0.594	0.468
Kurtosis	3.868	4.192	2.746	3.509	3.043	3.698
Jarque–Bera	176.120***	152.062***	106.372***	91.298***	219.376***	211.625***
ρ_1	0.949***	0.944***	0.964***	0.961***	0.960***	0.956***
ρ_{10}	0.778***	0.777***	0.857***	0.853***	0.836***	0.832***
ρ_{15}	0.701***	0.699***	0.824***	0.823***	0.792***	0.789***
ADF	−3.167**	−3.129**	−3.070**	−3.077**	−4.410***	−4.060***
PP	−4.964***	−5.129***	−5.389***	−5.571***	−7.477***	−7.880***
KPSS	1.192***	1.277***	1.477***	1.408***	0.767***	0.720**

The table reports summary statistics for the model-free implied correlation (MFIC) and the inverse logistic transformation (transformed MFIC). In addition, we also report the autocorrelation coefficient (ρ) for the 1st, 10th and 15th lags of the autocorrelation structure, as well as the augmented Dickey–Fuller (ADF), Phillips–Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test statistics. The number of lags for the ADF test is selected using the modified Schwarz criterion. The bandwidth for the PP and KPSS tests is selected automatically using the Newey–West lag selection parameter. The null hypothesis of the ADF and PP tests is the presence of a unit root, whereas the null hypothesis of the KPSS test is stationarity. The results are presented for two sub-periods: the first extends from January 3, 1996, to December 30, 2000 (columns 2–3), and the second runs from January 3, 2001, to October 29, 2010 (columns 4–5). We also report the results for the entire sample period (columns 6–7).

* Denote rejections of the null hypothesis at the 10% level.
** Denote rejections of the null hypothesis at the 5% level.
*** Denote rejections of the null hypothesis at the 1% level.

Fig. 1 shows the daily evolution of the model-free implied correlation from January 1996 to October 2010. Consistent with the well-documented fact of increased correlations during bear markets, the series present several spikes over the sample period, coinciding with periods of low returns and high volatilities. Specifically, as a result of the Asian and Russian financial crises, the MFIC series reached a value of 0.9 in late 1997, and again in August 1998. During the early years of the 21st century, the low returns market attributed to the 9/11 terrorist attack, the South-American financial turmoil and the internet bubble burst resulted in increased levels of correlation, although they were less notable than the aforementioned spikes in the 1990s. In subsequent years, the MFIC trended lower, with the exception of a period in May 2006, in which a global sell-off occurred. Finally, as a result of the global financial crunch, which commenced in 2007 and culminated in September 2008 with the announcement of the bankruptcy of Lehman Brothers, the MFIC presents several peaks, reaching its highest value of 0.97 on October 24, 2008.

Table 1 reports the summary statistics of the model-free implied correlation series for the in-sample, out-of-sample and full sample periods. The higher moments and the Jarque–Bera test reject the hypothesis of normality for the MFIC series for all of the periods under consideration. Table 1 also reports the test statistics from the augmented Dickey–Fuller (ADF), Phillips–Perron (PP), and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. According to the PP and ADF tests, the null hypothesis of a unit root is rejected strongly for all three subsets at a 1% significance level. Conversely, the KPSS test provides evidence of non-stationarity of the time series for all three periods under consideration. These mixed results, along with the slowly decaying and significant autocorrelation coefficients, suggest the existence of long memory properties in the MFIC, similar to the model-free implied volatility (see Brooks & Oozeer, 2002). We discuss these features further in the next section.

2. Forecasting models

By construction, for a sufficiently large N, the MFIC is bounded by 0 and 1 (Bourgoin, 2001). Moreover, a necessary condition for the positive definiteness of a correlation matrix with all off-diagonal elements equal to $\bar{\rho}_t$ is that $\bar{\rho}_t$ lies within $[-1/(n-1), 1]$ (see Engle & Kelly, 2012). We examine the predictability of the MFIC series and alleviate any issues that may arise in the forecasting procedure by using the inverse of the logit transformation of the original correlation series, defined as

$$\rho_t^* = -\frac{1}{n-1} + \left(1 + \frac{1}{n-1}\right) \frac{1}{1 + \exp(-\rho_t)}, \tag{4}$$

where ρ_t^* is the model-free implied correlation series, n is the sample size, and ρ_t is the logit transformation of the series. The logit transformation ensures that the series will lie in the interval $[-1/(n-1), 1]$.

Table 1 also presents the summary statistics for the logit transformation of the series. Unit root (ADF and PP) and stationarity (KPSS) tests provide mixed results regarding the stationarity of the logit transformation, similar to those obtained for the original MFIC series. We investigate the stationarity and long memory properties of our series further by estimating an ARFIMA(p, d, q) model for the logit transformation. The fractional differencing parameter is significant and equal to 0.5, thus confirming the non-stationarity of our series. Therefore, the remainder of the paper applies the proposed econometric specifications to the differentiated logit transformation of the MFIC series. However, for reasons of notation, we shall refer to the first difference in the logit transformation as the MFIC series.

Stock returns and implied volatility have been examined extensively for the presence of seasonality effects. Monday (Friday) tends to be a day on which traders open (close) positions for the week, and the excess buying (selling) pressure may result in a higher (lower) implied volatility (see Harvey & Whaley, 1992). Therefore, we investigate the presence of day-of-the-week effects using the following specification:

$$MFIC_t = \sum_{i=1}^{5} \gamma_i D_{i,t} + \delta MFIC_{t-1} + u_t, \tag{5}$$

where i takes values of 1 to 5 for Monday to Friday, respectively.

In addition, the following specification is employed for assessing the significance of the January effect in the evolution of the MFIC:

$$MFIC_t = \sum_{i=1}^{12} \gamma_i D_{i,t} + \delta MFIC_{t-1} + u_t, \tag{6}$$

where i takes values of 1 to 12 for January to December, respectively. The inclusion of a lagged term of the dependent variable in both Eqs. (5) and (6) accounts for the elimination of autocorrelated error terms.

A number of econometric specifications that capture different aspects of the distributional characteristics are applied to model the evolution of the series and ultimately to assess the predictability of the model-free implied correlation series. Next, the specifications are described thoroughly.

2.1. AR(I)MA, AR(I)MA-GARCH and ARFIMA models

First, the MFIC is modeled using an autoregressive moving average (AR(I)MA) process by adding both the lags of the error term and the lags of the series under examination. The AR(I)MA(r, d, m) model can be specified by:

$$\Phi(L)\Delta^d (X_t - \mu) = \Theta(L) u_t, \tag{7}$$

where Φ and Θ are polynomials of orders r and m, respectively, such that $\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_r L^r$, and $\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_m L^m$, X_t is the MFIC, μ is the expectation of X_t, ε_t is the white noise error term, Δ is the difference operator and d is equal to one, representing the order of integration.

Furthermore, we also employ an AR(I)MA(r, d, m)-GARCH(p, q) model that accounts for the remaining heteroskedasticity in the error term. The error term is assumed to follow a normal distribution, with a zero mean and variance equal to h_t^2, where:

$$h_t^2 = b_0 + \sum_{q=1}^{q} b_q u_{t-q}^2 + \sum_{p=1}^{p} c_p h_{t-p}^2. \tag{8}$$

Taking the AR(I)MA model one step further by allowing d to take fractional values, the specification extends to an ARFIMA(r, d, m) model, given by:

$$\Phi(L)(1-L)^d (X_t - \mu) = \Theta(L) u_t, \tag{9}$$

where $(1-L)^d$ is the fractional integration operator. The ARFIMA model captures both the short-run component of the series, through the autoregressive and moving average terms, and the long-run dependence, through the fractional differencing parameter d. The process is stationary under the condition that $-0.5 < d < 0.5$. If $0 < d < 0.5$, the series exhibits long memory, suggesting positive dependence between distant observations, whereas if $-0.5 < d < 0$, the series presents negative dependence between distant observations, known as anti-persistence. Finally, for $d = 0$ (the general AR(I)MA process), the process exhibits short memory.

The AR(I)MA, AR(I)MA-GARCH and ARFIMA models have been estimated for all plausible combinations of autoregressive (AR) and moving average (MA) order terms, in both the mean and variance specifications, where applicable, up to the fifth lag. The model selection is based on the Schwarz information criterion, which, unlike the Akaike information criterion, includes an extra term that penalizes the overfitting of data.

2.2. Regime-switching model

In addition, we employ a dynamic regime-switching model for capturing potential asymmetries in the correlation process. More specifically, the transition between regimes is governed by a Markov chain, two regimes are assumed, and the coefficient on the lagged dependent variable is allowed to be regime-varying, i.e.,

$$\sum_{r=1}^{R} \Phi_{s_t}^r (L)\, \Delta^d \left(X_t - \mu_{s_t}^r\right) = u_t. \tag{10}$$

The transitions between regimes $s_t = 1$ and $s_t = 2$ are given by a Markov chain with transition probabilities

$p_{ij} = P(s_t = j|s_{t-1} = i)$ for $i, j = 1, 2$ and $\sum_{j=1}^{2} p_{ij} = 1$ for $i = 1, 2$. The lag order is selected based on the Schwarz information criterion.

2.3. Heterogeneous autoregressive model

The heterogeneous market hypothesis is based on the empirical finding that the volatility dynamics are affected by the differences in traders' investment horizons. Specifically, traders with short-term investment horizons incorporate any new information into their strategy rapidly, thus affecting the shorter-term volatility directly. Conversely, longer-term investors rebalance their positions less frequently and disregard daily information flows, thus empowering the long-memory characteristic of the volatility. Corsi (2009) identified and modeled this asymmetry through an autoregressive process by aggregating daily, weekly and monthly volatilities.

Given the superior forecasting performance of the heterogeneous autoregressive (HAR) model that has been documented in the volatility context, we extend its usage in the correlation context and examine its ability to gauge the long-memory property of the model-free implied correlation. The model is given by:

$$MFIC_t^{(d)} = \alpha_0 + \alpha_{(d)} MFIC_{t-1}^{(d)} + \alpha_{(w)} MFIC_{t-1}^{(w)} + \alpha_{(m)} MFIC_{t-1}^{(m)} + u_t, \quad (11)$$

where $MFIC_{t-1}^{(w)} = \frac{1}{5}(\sum_{h=1}^{5} MFIC_{t-h}^{(d)})$ is the lagged weekly MFIC, and $MFIC_{t-1}^{(m)} = \frac{1}{22}(\sum_{h=1}^{22} MFIC_{t-h}^{(d)})$ is the lagged monthly MFIC.

2.4. Economic determinants

Several studies have examined the roles of economic variables in determining asset return correlations (see Erb, Harvey, & Viskanta, 1994; Moskowitz, 2003; Palandri, 2009; and Sheppard, 2008; amongst others). These studies indicate that financial and macroeconomic variables such as short-term interest rates or the slope of the term structure would be expected to affect the systematic risk of equity portfolios and the time variation of equity correlations. Data for all explanatory variables are obtained from Datastream.

In line with Harvey and Whaley (1992) and Sheppard (2008), we include three interest rate variables that predictability studies have shown broadly to influence the volatility and correlation process: the one-month USD LIBOR interest rate (INT_t); the slope of the yield curve (or else the term structure, $TERM_t$), defined as the difference between the yield of the 10-year US government bond and the yield of the 3-month Treasury bill; and the slope of the junk bond spread ($JUNK_t$), defined as the difference between an Aaa-rated bond yield and a Baa-rated bond yield. The short-term interest rate is considered to proxy the shocks to the expected real economic activity accurately. The junk bond spread and the term structure have been documented previously to capture the long- and short-term business cycle conditions, respectively. The lagged return and the trading volume (VOL_t) of the underlying security are included as control variables for leverage and information flow effects (see Bollen & Whaley, 2004). An increase in trading volume signals to investors the arrival of new information, thus inducing fluctuations in both returns and implied volatilities. We also include two dummy variables for assessing the asymmetric response of the MFIC to positive (R_t^+) and negative (R_t^-) index returns. In addition, the West Texas intermediate price (WTI_t) has also been considered as a proxy for fluctuations in an alternative asset class market. Further explanatory variables include the dividend yield of the S&P 100 index (DIV_t), which is used as a proxy for the time-varying expected returns, and the EURO/USD exchange rate (FX_t). Each of these variables is differentiated in order to represent innovations and ensure stationarity.

Before proceeding to the estimation, the model was tested for inference issues that can arise when regressing returns and for volatility in the macroeconomic variables (see Paye, 2008). In order to avoid multicollinearity issues based on the correlation matrix of the dependent variables, we do not include the MFIV changes in the index. Second, we use Granger causality tests to test for endogeneity in the model based on possible reverse causality between the dependent and independent variables. The only dependent variable that exhibits reverse causality with the dependent variable is the trading volume. However, the contemporaneous correlation between the MFIC and the trading volume is not significant. In order to minimize any bias in the forecasted values, we have used the lagged values of the macroeconomic variables.[6] Finally, the estimated model includes up to three lags in the MFIC series to account for autocorrelation patterns, and the model that minimizes the Schwarz information criterion is selected as follows:

$$\begin{aligned} MFIC_t = {} & b_0 + b_1^+ R_{t-1}^+ + b_2^- R_{t-1}^- + b_3 FX_{t-1} \\ & + b_4 INT_{t-1} + b_5 DIV_{t-1} + b_6 TERM_{t-1} \\ & + b_7 JUNK_{t-1} + b_8 WTI_{t-1} \\ & + b_9 VOL_{t-1} + u_t. \end{aligned} \quad (12)$$

Table 2 presents the descriptive statistics for the economic determinants. Although stationarity is rejected when the variables are measured in levels (Panel A), unit root tests conducted on the first difference indicate that all series are stationary (Panel B). We record significant first order autocorrelation coefficients for all variables measured in levels, but find no evidence of significant autocorrelation in the majority of the cases when the variables are measured in first differences.

2.5. Combination forecasts

Each forecasting model captures different dynamics regarding the underlying asset properties, based on alternative information sets. The empirical results from previous studies (see Becker & Clements, 2008) suggest that the forecasting performances of linear combinations

6 The results of preliminary correlation analyses and Granger causality tests are available from the authors upon request.

Table 2
Descriptive statistics for the economic determinants.

	Index closing price	DIV	Volume (*VOL*)	FX	JUNK	Libor (*r*)	TERM	WTI
Panel A: raw series								
Mean	544.850	0.015	864,752,943.520	1.103	0.007	0.057	0.009	21.361
Median	531.970	0.014	860,160,000.000	1.103	0.007	0.056	0.009	20.680
Maximum	832.650	0.022	2,405,100,000.000	1.312	0.011	0.068	0.019	37.220
Minimum	285.770	0.010	136,580,000.000	0.825	0.005	0.049	−0.007	10.820
Std. Dev.	166.758	0.004	208,760,374.440	0.118	0.001	0.005	0.006	5.885
Skewness	0.088	0.330	0.380	−0.258	0.956	0.820	−0.464	0.420
Kurtosis	1.636	1.731	5.313	2.437	3.437	3.111	2.592	2.554
Jarque–Bera	99.091***	107.097***	310.450***	30.532***	201.575***	141.573***	53.770***	47.419***
ρ_1	0.998***	0.997***	0.727***	0.997***	0.993***	0.994***	0.991***	0.995***
ADF test	−1.272	−1.487	−3.087	−1.160	−1.879	−1.108	−0.713	−1.644
Panel B: first difference of the series								
Mean	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.001
Maximum	0.056	0.001	2.031	0.042	0.001	0.144	0.005	0.145
Minimum	−0.075	−0.001	−1.917	−0.023	−0.001	−0.108	−0.004	−0.150
Std. Dev.	0.012	0.000	0.221	0.006	0.000	0.008	0.001	0.026
Skewness	−0.331	−0.020	0.290	0.615	−0.750	2.713	0.859	−0.148
Kurtosis	6.299	20.352	27.379	5.781	11.537	183.555	12.764	7.172
Jarque–Bera	592.630***	15,757.618***	31,121.145***	483.816***	3,931.412***	1,707,612.708***	5,143.726***	915.461***
ρ_1	−0.018	−0.002	−0.360***	−0.039	−0.056**	0.027	0.021***	−0.002
ADF test	−95.973***	−35.479***	−49.516***	−6.664***	−6.523***	−7.108***	−34.683	−35.237***

The table reports the in-sample summary statistics for the variables employed in the economic determinants model. The index closing price, *DIV* and *VOL* refer to the closing price, the dividend yield and the trading volume of the S&P 100 index, correspondingly. *FX* is the EUR/USD exchange rate, *JUNK* is the slope of the junk bond spread, defined as the difference between an Aaa bond yield and a Baa bond yield, *Libor (r)* is the one-month USD LIBOR interest rate, *TERM* is the slope of the yield curve (or the term structure), defined as the yield of the 10-year US government bond minus the yield of the 3-month Treasury bill, and *WTI* is the price of Brent crude oil. In addition, we also report the autocorrelation coefficient (ρ) for the first lag of the autocorrelation structure and the augmented Dickey–Fuller (ADF) test statistics. The number of lags for the ADF test is selected using the modified Schwarz criterion. The null hypotheses of the ADF and Jarque–Bera tests are the presence of the unit and normal distributions, respectively.

* Denote rejections of the null hypothesis at the 10% level.
** Denote rejections of the null hypothesis at the 5% level.
*** Denote rejections of the null hypothesis at the 1% level.

of individual forecasts may be superior to those of the individual forecasts. In this section, we present and perform a comparative analysis of the various methods proposed in the literature for combining forecasts.

At time t, we obtain six different forecasted values of the MFIC from the time series models described above (denoted by f_t^j, where $j = 1$ for AR(I)MA, 2 for AR(I)MA-GARCH, 3 for ARFIMA, 4 for the regime-switching model, 5 for the HAR and 6 for the economic determinants model). For notational reasons, we use the same superscripts for the forecasting models throughout our descriptions of the methods below.

The simplest method of combining forecasts is to assign an equal weight to each individual forecast. Thus, the equal combination forecast is a simple average of the individual forecasts, as follows:

$$f_t^{EW} = \frac{1}{6}\sum_{j=1}^{6} f_t^j. \tag{13}$$

In addition, the existing literature proposes the usage of the Schwarz model selection criterion as an alternative approach for obtaining combination weights. Toward this end, each model's weight is calculated as the difference between its Schwarz criterion value and that of the "best" performing model that obtains the minimum value of the criterion. Using the methodology of Kolassa (2011), the Schwarz weighted combination forecast equals:

$$f_t^{BIC} = \sum_{j=1}^{6} w_{BIC}(j) \cdot f_t^j, \tag{14}$$

where $w_{BIC}(j) = \frac{\exp(-\frac{1}{2}\Delta_{BIC}(j))}{\sum_{j=1}^{6}\exp(-\frac{1}{2}\Delta_{BIC}(j))}$, $\Delta_{BIC}(j) = BIC(j) - BIC(k)$, $BIC(j)$ is the Schwarz criterion of the jth model, $j = 1, 2, \ldots, 6$, and $BIC(k)$ is the minimum Schwarz criterion of model k.

The aforementioned combination forecasts consider constant weights throughout the out-of-sample period, and therefore fail to capture the dynamics of the series under consideration. To overcome this restriction, we also consider the derivation of time-varying weighted combination forecasts. In essence, the weights are obtained by minimizing the mean squared error of the regression:

$$MFIC_t = \beta_0 + \beta_1 f_{t|t-1}^1 + \beta_2 f_{t|t-1}^2 + \beta_3 f_{t|t-1}^3 + \beta_4 f_{t|t-1}^4 + \beta_5 f_{t|t-1}^5 + \beta_6 f_{t|t-1}^6, \tag{15}$$

where $MFIC_t$ is the actual value of the model-free implied correlation series and $f_{t|t-1}^j$ are the forecasted values of individual models at time t, calculated at time $t-1$. As a consequence, the combination forecast value is given by:

$$f_{t+1|t}^w = \hat{\beta}_0 + \hat{\beta}_1 f_{t+1|t}^1 + \hat{\beta}_2 f_{t+2|t}^2 + \hat{\beta}_3 f_{t+3|t}^3 + \hat{\beta}_4 f_{t+4|t}^4 + \hat{\beta}_5 f_{t+5|t}^5 + \hat{\beta}_6 f_{t+6|t}^6. \tag{16}$$

Table 3
Intra-week pattern.

	Coefficient	t-statistic
Monday	0.082***	6.999
Tuesday	0.010	0.886
Wednesday	−0.015	−1.610
Thursday	−0.005	−0.357
Friday	−0.067***	−5.426
$MFIC_{t-1}$	−0.203***	−4.819

The table reports the coefficients and t-statistics estimated from the model described in Eq. (6). The reported t-statistics have been calculated using Newey–West autocorrelation consistent standard errors, with the number of lags selected automatically according to the Schwarz criterion.
* Indicate rejection of the null hypothesis at the 10% significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

Table 4
Monthly seasonality.

	Coefficient	t-statistic
January	−0.001	−0.036
February	−0.002	−0.140
March	0.013	0.831
April	−0.011	−0.732
May	0.005	0.433
June	−0.015	−1.222
July	0.022	1.279
August	0.004	0.264
September	−0.003	−0.198
October	0.006	0.260
November	−0.012	−0.695
December	−0.005	−0.292
$MFIC_{t-1}$	−0.222***	−4.811

The table reports the coefficients and t-statistics estimated from the model described in Eq. (5). The reported t-statistics have been calculated using Newey–West autocorrelation consistent standard errors, with the number of lags selected automatically according to the Schwarz criterion.
* Indicate rejection of the null hypothesis at the 10%, significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

The implementation of this two-step approach for obtaining weighted combination forecasts over the out-of-sample period, extending from January 2001 to October 2010, is conducted as follows. First, data covering the period from January 1996 to June 1998 are used to estimate the various models under consideration. Next, we obtain one-step-ahead rolling forecasts over the "pseudo" out-of-sample period, commencing in July 1998 and ending in December 2000. We then use these forecasts to obtain the weights by estimating the regression specification described in Eq. (15).

Following the methodology of Aksu and Gunter (1992), we estimate Eq. (15) using four alternative approaches, by imposing alternative restrictions on the coefficients and the constant term of the regression. As an initial test, the constant term is excluded and no restrictions are applied to the values of the weights (Method A). Next, the sum of the weights of alternative forecasts is restricted to unity and no constant term is included (Method B). An alternative specification of Eq. (15) includes a constant term, but with no restriction applied to the value of the weights (Method C). Finally, we estimate the equation with a constant term while restricting the weights to have positive values and to sum to unity (Method D).

Table 5
In-sample evidence from the AR(I)MA, ARFIMA and AR(I)MA-GARCH models.

	AR(I)MA	ARFIMA	AR(I)MA-GARCH
Constant	−0.008*	−0.008*	−0.012***
d	–	0.000	–
φ_1	0.685***	0.901***	0.760***
φ_2	0.118**	–	−0.980***
φ_3	–	–	0.771***
θ_1	−0.934***	−1.144***	−0.956***
θ_2	–	0.139***	1.075***
θ_3	–	0.027	−0.992***
θ_4	–	–	0.075***
θ_5	–	–	−0.040***
b_0	–	–	0.003***
b_1	–	–	0.075***
b_2	–	–	0.291***
b_3	–	–	0.197***
b_4	–	–	−0.211**
b_5	–	–	−0.193***
c_1	–	–	−0.999***
c_2	–	–	0.939***
c_3	–	–	0.887***
c_4	–	–	−0.065***
Monday dummy	0.101***	0.100***	0.104***
Friday dummy	−0.053***	−0.055***	−0.046***
Log likelihood	293.341	299.442	465.785
Adj. R^2	0.122	0.132	0.125
BIC	−0.434	−0.426	−0.624
$Q^2(20)$	258.043***	237.270***	13.300
ARCH test (5)	154.488***	146.620***	2.571

Column 1 reports the estimates from an AR(I)MA (2,1) model using the specification: $\Phi(L)\Delta^d(MFIC_t - \mu) = \Theta(L)u_t$. Column 2 refers to the estimates obtained from an ARFIMA (1, d, 3) model. The specification employed is: $\Phi(L)(1-L)^d(MFIC_t-\mu) = \Theta(L)u_t$. The last column presents the estimates from the AR(I)MA(3,5)–GARCH(5,4) model, specified as $\Phi(L)\Delta^d(MFIC_t - \mu) = \Theta(L)u_t$. The significant day-of-the-week variables reported earlier are also included in each specification. The t-statistics for the AR(I)MA model are estimated using the HAC estimates of standard errors. The log-likelihood value, the adjusted R^2, the value of the Schwarz criterion, the Ljung–Box statistic for the 20th lag of squared residuals and the F-statistic (fifth lag) for remaining heteroskedasticity are also reported.
* Indicate rejection of the null hypothesis at the 10% significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

In the final step, the out-of-sample weighted combination forecasts are derived by substituting the previously estimated weights and the out-of-sample forecasted values of individual models into Eq. (16).

3. In-sample estimation results

The in-sample results suggest that the MFIC series clearly present a strong intra-week pattern, because the Monday and Friday dummy variables are strongly significant, whereas there is no evidence of monthly seasonality for the MFIC series (Tables 3 and 4, respectively). As expected, the Monday (Friday) dummy variable has a significant positive (negative) coefficient, which is consistent with the reported increase in buying (selling) activity on these specific days of the week.

Table 5 presents the in-sample estimation results from the AR(I)MA, ARFIMA and AR(I)MA-GARCH models, while

Table 6
In-sample evidence from the regime-switching model.

	Coefficient	t-statistic
μ_1	0.870***	11.967
ϕ_1^1	−1.513***	−17.239
ϕ_1^2	0.868***	3.740
ϕ_1^3	1.161***	6.141
ϕ_1^4	0.728***	4.935
μ_2	−0.008	−1.367
ϕ_2^1	−0.142***	−5.212
ϕ_2^2	−0.081***	−3.174
ϕ_2^3	−0.091***	−3.535
ϕ_2^4	−0.011	−0.425
log(σ)	−1.791***	−85.421
Monday dummy	0.082***	6.421
Friday dummy	−0.057***	−4.505
Log likelihood	418.623	
Adj. R^2	−0.079	
BIC	−0.583	
$Q^2(20)$	527.506***	
ARCH test (5)	377.959***	

The table presents the estimated coefficients and the associated t-statistics from the regime-switching model, estimated using the following specification: $\sum_{r=1}^{R} \Phi_{S_t}^r (L)\, \Delta^d \left(X_t - \mu_{S_t}^r\right) = u_t$. The significant day-of-the-week variables reported earlier are also included in the specification. The log-likelihood value, the adjusted R^2, the value of the Schwarz criterion, the Ljung–Box statistic for the 20th lag of squared residuals and the F-statistic (fifth lag) for remaining heteroskedasticity are also reported.

* Indicate rejection of the null hypothesis at the 10% significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

Table 7
In-sample evidence from the heterogeneous autoregressive (HAR) model.

	Coefficient	t-statistic
α_0	−0.024***	−3.430
$\alpha_{(d)}$	−0.129**	−2.534
$\alpha_{(w)}$	−0.307**	−2.456
$\alpha_{(m)}$	−0.902***	−2.694
Monday dummy	0.017	1.296
Friday dummy	0.108***	7.774
Log likelihood	276.132	
Adj. R^2	0.109	
BIC	−0.413	
$Q^2(20)$	264.281***	
ARCH test (5)	163.561***	

The table presents the estimated coefficients and the associated t-statistics from the HAR model, estimated using the specification $MFIC_t^{(d)} = \alpha_0 + \alpha_{(d)} MFIC_{t-1}^{(d)} + \alpha_{(w)} MFIC_{t-1}^{(w)} + \alpha_{(m)} MFIC_{t-1}^{(m)} + u_t$. The significant day-of-the-week variables reported earlier are also included in the specification. The t-statistics of the model have been estimated using the HAC estimates of standard errors. The log-likelihood value, the adjusted R^2, the value of the Schwarz criterion, the Ljung–Box statistic for the 20th lag of squared residuals and the F-statistic (fifth lag) for remaining heteroskedasticity are also reported.

* Indicate rejection of the null hypothesis at the 10% significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

Tables 6–8 present those for the regime-switching, heterogeneous autoregressive (HAR) and economic determinants models, respectively. The AR(I)MA(2, 1), ARFIMA(1, d, 3), AR(I)MA(3, 5)-GARCH(5, 4) and regime-switching (4, 0) models were found to minimize the Schwarz criterion.

Table 8
In-sample evidence from the economic determinants model.

	Coefficient	t-statistic
b_0	0.017	1.498
R_{t-1}^-	3.336***	3.391
R_{t-1}^+	−1.452	−1.405
FX_{t-1}	−1.055	−1.321
r_{t-1}	−0.038	−0.049
DIV_{t-1}	8.983	0.211
$TERM_{t-1}$	−11.194	−1.415
$JUNK_{t-1}$	−80.318*	−1.932
WTI_{t-1}	−0.033	−0.196
VOL_{t-1}	−0.050*	−1.714
$MFIC_{t-1}$	−0.181***	−3.322
$MFIC_{t-2}$	−0.088**	−2.246
$MFIC_{t-3}$	−0.007	−0.123
Monday dummy	0.084***	5.705
Friday dummy	−0.062***	−4.067
Log likelihood	298.160	
Adj. R^2	0.122	
BIC	−0.391	
$Q^2(20)$	222.427***	
ARCH test (5)	117.069***	

The table presents the estimated coefficients and the associated t-statistics from the economic determinants model, estimated using the specification: $MFIC_t = b_0 + b_1^+ R_{t-1}^+ + b_2^- R_{t-1}^- + b_3 FX_{t-1} + b_4 r_{t-1} + b_5 DIV_{t-1} + b_6 TERM_{t-1} + b_7 JUNK_{t-1} + b_8 WTI_{t-1} + b_9 VOL_{t-1} + b_{10} MFIC_{t-1} + b_{11} MFIC_{t-2} + b_{12} MFIC_{t-3} + u_t$, where R_{t-1}^+ and R_{t-1}^- represent the positive and negative log-index returns, FX_{t-1} is the log-difference of the EUR/USD exchange rate, r_{t-1} is the log-difference of the one-month USD LIBOR interest rate, DIV_{t-1} is the change in the dividend yield of the underlying index, $TERM_{t-1}$ is the slope of the yield curve (or the term structure), defined as the yield of the 10-year US government bond minus the yield of the 3-month Treasury bill, $JUNK_{t-1}$ is the slope of the junk bond spread, defined as the difference between an Aaa bond yield and a Baa bond yield, WTI_{t-1} is the log-difference of Brent crude oil, and VOL_{t-1} is the log-difference of the trading volume of the S&P 100 index. We also include two lagged values of the MFIC series. The significant day-of-the-week variables reported earlier are also included in the specification. The t-statistics of the model have been estimated using the HAC estimates of standard errors. The log-likelihood value, the adjusted R^2, the value of the Schwarz criterion, the Ljung–Box statistic for the 20th lag of squared residuals and the F-statistic (fifth lag) for remaining heteroskedasticity are also reported.

* Indicate rejection of the null hypothesis at the 10% significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

Several diagnostic tests were employed for testing the goodness-of-fit of the various models. We report the log likelihood value, the adjusted R^2, the Schwarz criterion, the Ljung–Box statistic for the 20th lag of squared residuals, and the F-statistic of the ARCH test for the remaining heteroskedasticity up to the fifth lag. The inclusion of the GARCH specification in the residuals corrects for the remaining serial autocorrelation and heteroskedasticity. The AR(I)MA-GARCH specification produces the lowest value of the Schwarz criterion and the maximum log-likelihood value, whereas the ARFIMA model produces the highest value of R^2. The null hypotheses of both no serial autocorrelation and no heteroskedasticity in the residual terms are rejected strongly for the AR(I)MA and ARFIMA models. We also note that the value of the fractional differencing parameter d in the ARFIMA specification is significant, suggesting that the series does not exhibit long memory features; a simple ARMA(1, 3) model is able to capture the dynamics of the series.

Table 9
Evaluation of out-of-sample performances.

Panel A										
	AR(I)MA	AR(I)MA-GARCH	ARFIMA	RS	HAR	ED	Equal-weighted combination	Time-varying weighted combination	Schwarz-weighted combination	Random walk
RMSE	17.828%	17.905%	18.010%	17.976%	17.943%	17.960%	17.716%	17.907%	17.715%	18.578%
MAE	11.851%	11.882%	12.132%	12.005%	12.125%	12.004%	11.804%	11.891%	11.803%	12.468%
MSE-F	212.427***	189.389***	158.358***	168.501***	177.996***	173.101***	246.508***	188.671***	246.676***	

Panel B									
	AR(I)MA	AR(I)MA-GARCH	ARFIMA	RS	HAR	ED	Equal-weighted combination	Time-varying weighted combination	Schwarz-weighted combination
MCP	59.773%	59.976%	56.860%	59.409%	57.345%	58.357%	60.542%	59.733%	60.542%
PT test	9.585***	9.728***	6.587***	9.199***	7.452***	8.208***	10.372***	9.640***	10.369***

Panel A presents the root mean square error (RMSE) and the mean absolute error (MAE) in percentage terms, corresponding to the AR(I)MA, AR(I)MA-GARCH, ARFIMA, regime-switching (RS), HAR, economic determinants (ED) and random walk models, as well as the combination forecasts obtained with equal, time-varying and Schwarz weights. The McCracken F-test for the MSE is used to test the null hypothesis that the model produces a forecasting accuracy that is equal to that of the random walk model, against the alternative of the former being more accurate. Panel B presents the directional forecasting accuracy of the models under consideration. The mean correct prediction (MCP) measure represents the number of times that the actual change in the MFIC is predicted correctly by the forecasted series. The PT test is employed to measure the independence between the actual and forecasted values.

* Indicate rejection of the null hypothesis at the 10% significance level.

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

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

Table 6 presents the results of the estimated regime-switching model. The results for the two regimes differ significantly. The conditional mean term in the first regime is significantly different from zero and equal to 0.87, whereas the conditional mean in the second regime is not significantly different from zero. The first regime is characterized by high MFIC changes, whereas the second regime is characterized by low MFIC changes and negative mean reversion. Moreover, the persistence is high in the first regime ($p_{22} = 0.99$), but significantly lower in the second regime ($p_{11} = 0.46$).

Table 7 presents the results from the heterogeneous autoregressive model. The coefficients of all three estimates of the implied correlation, corresponding to different time horizons, are all highly significant and negative. We observe that the impact of the lagged correlation strengthens as the time horizon of aggregation increases.

Table 8 depicts the estimation results from the economic determinants model with three lagged terms of the MFIC, as dictated by the minimum Schwarz criterion. Interestingly, the only variables that are statistically significant at the 10% level are the first and second lagged values of the MFIC, the lagged negative index returns, the lagged junk spread, the lagged volume of the S&P 100 index, and the dummy variables representing the Monday and Friday effects. Moreover, unreported results confirm the asymmetric response of the MFIC series to negative and positive contemporaneous index returns. The negative coefficients of the lagged default spread and the trading volume suggest that an increase in the variables at time $t-1$ will actually reduce the current level of the MFIC series. This effect of the variables decreasing the series, together with the divergence from the widely-reported positive contemporaneous relationship between the above-mentioned variables and the volatility, might stem from the information flow and the attention shift hypothesis, as proposed by Peng (2005) and Peng and Xiong (2006). The fundamental underlying concept of the hypothesis lies in investors' selective information processing, based on priority and urgency. Specifically, taking into account the fact that the market factor carries more weight in most portfolios, an increase in the number of variables under consideration will cause a direct increase in the market uncertainty, and investors will focus on resolving the market uncertainty before any other asset-specific information, resulting in an increase in the contemporaneous stock market correlations. Peng, Xiong, and Bollerslev (2007) examined the effect of this hypothesis on stock return co-movements and found that macroeconomic shocks lead to a contemporaneous increase in the co-movement of individual stocks with the market. However, in the subsequent period, investors shift their attention to asset-specific news, meaning that equity correlations are expected to decrease in the following days.

One-step-ahead forecasts are obtained from the estimation of the economic determinants model using only the variables that are significant at the 10% significance level.

4. Evaluation of out-of-sample forecasting performances

4.1. Statistical measures

In this section, we present the results and assess the predictive abilities of the out-of-sample forecasts constructed using the models discussed above. For the first out-of-sample forecast, corresponding to January 2, 2001, we estimate the models for the sample period from January 4, 1996, to December 29, 2000, and obtain the relevant forecasted value for the following day. We then use a rolling window sample to produce one-step-ahead forecasting values for the entire out-of-sample period.

Panel A of Table 9 depicts the performances of the AR(I)MA, AR(I)MA-GARCH, ARFIMA, regime-switching, HAR and economic determinants models, with the last four columns demonstrating the results from the alternative combination methods and the random walk model. The equal- and Schwarz-weighted combination forecasts were obtained from Eqs. (13) and (14), respectively. Combination forecasts using time-varying weights have been formulated via the four alternative approaches described in Section 2. Consistent with the findings of Aksu and Gunter (1992), the forecasts obtained from Eq. (15), estimated using a constant term and under the restriction of positive weights that sum to unity, minimize the RMSE and are chosen to compete with the other specifications thereafter. For the remainder of the paper, we shall refer to the results and the forecasting performance of the weighted combination forecast from Method D as the time-varying weighted combination forecast.

The out-of-sample forecasting performances of the models presented above and the predictability of the model-free implied correlation series are initially assessed using statistical measures. The root mean squared error (RMSE) and the mean absolute error (MAE) are used to evaluate the magnitudes of the forecast errors. The RMSE is calculated as the square root of the average squared deviations of the forecasted values from the actual series, whereas the MAE is measured as the average of the absolute values of the forecast errors. Notably, neither the RMSE nor MAE metrics vary substantially across the models employed. In terms of both the RMSE and the MAE, the minimum values are attributed to the Schwarz-weighted combination model, being only 0.0005% and 0.0012% lower than the respective values of the equal combination forecast. Our results support existing studies in suggesting that combination forecasts, which essentially accumulate the forecasts of several econometric models, yield forecasting performances that are superior to those of single models.

The comparison of the forecasting accuracies of the individual models with that of the random walk benchmark constitutes a direct test for the presence of market efficiency. Taking into account the fact that the random walk model is nested to all alternative models, we employ the McCracken (2007) test statistic, which is valid for the comparison of nested models. The F-type statistic is defined as below:

$$MSE - F = (T - h + 1)\frac{MSE_N - MSE_A}{MSE_A}, \tag{17}$$

where T is the number of out-of-sample forecasts, $h = 1$ is the h-step-ahead forecasts, MSE_N is the MSE of the nested model (i.e., the random walk model) and MSE_A is the MSE of the alternative models. According to the null hypothesis, the two models have equal MSEs, whereas the alternative is that the MSE of the nested model is larger than that of the alternative one. The asymptotic distribution of the test under the null hypothesis is non-standard. McCracken (2007) provides asymptotically valid critical values according to the in-sample and out-of-sample observations and the numbers of parameter restrictions. The McCracken F-statistics calculated are reported in Panel A of Table 9. The results suggest that the null hypothesis of an equal forecasting accuracy is rejected strongly, and propose that each alternative model produces a smaller forecasting error than the benchmark model.[7]

The above-mentioned tests address the question of forecasting accuracy in terms of correct predictions of the magnitude. However, practitioners (and traders in particular) are interested primarily in predicting the directional change of their portfolio correlation correctly, in order to take the appropriate positions. We begin by assessing the directional predictability of the various models employed, using the mean correct prediction (MCP) measure. The MCP is computed as the percentage of observations for which the forecasting model predicted the realized direction of change in the MFIC correctly (see Goncalves & Guidolin, 2006). Second, we employ the non-parametric market-timing (PT) test introduced by Pesaran and Timmermann (1992). For the purpose of the PT test, we create a contingency table of realized and forecasted values.

		Actual value (y_{t+1})	
		Up ($y_{t+1} = 1$)	Down ($y_{t+1} = 0$)
Forecasted value ($\hat{y}_{t+1}$)	Up ($\hat{y}_{t+1} = 1$)	Hits (N_{uu})	False alarms (N_{ud})
	Down ($\hat{y}_{t+1} = 0$)	Misses (N_{du})	Correct rejections (N_{dd})

The PT test statistic could be expressed as follows:

$$PT = \frac{\sqrt{T}(H - F)}{\sqrt{\frac{\hat{\pi}_f(1-\hat{\pi}_f)}{\hat{\pi}_\alpha(1-\hat{\pi}_\alpha)}}}, \tag{18}$$

where T is the sample size, $H = N_{uu}/(N_{uu} + N_{du})$ is the percentage of correctly predicted "up" moves, $F = N_{ud}/(N_{ud} + N_{dd})$ is the proportion of "false alarms", $\hat{\pi}_\alpha = (N_{uu} + N_{du})/T$ is the probability that the actual series will move upwards, and $\hat{\pi}_f = (N_{uu} + N_{ud})/T$ is the probability that the forecasted series will move upwards. The PT test follows the standard normal distribution using the null hypothesis of independence between the actual and forecasted values, i.e., the forecasted series are not able to predict the sign of the actual series.

Panel B of Table 9 reports the results from the tests of directional accuracy. Remarkably, the MCP values do not vary significantly among the competing models, with the equal and Schwartz combination forecasts predicting the direction of the forecasted value successfully 60.54% of the time within the out-of-sample dataset. The PT test suggests that the forecasted and actual series are not distributed independently in all cases, and that there is in fact a predictable pattern in the direction of changes in the MFIC series.

In addition, the pairwise forecasting accuracy of the alternative specifications is assessed using the McCracken F-test (as described in Eq. (17)) and the modified Diebold–Mariano test for nested and non-nested models, respectively.[8] We define the loss differential function $d_{ijt} = [g(e_{jt}) - g(e_{it})]$, $i \neq j$, where $g(e_{ijt})$ is the loss function, and $i, j = 1$ for AR(I)MA, 2 for AR(I)MA-GARCH, 3 for

[7] Unreported results from the t-type test proposed by McCracken (2007) consistently lead to the same conclusion.

[8] Note that the AR(I)MA is nested in the AR(I)MA-GARCH specification, and each individual forecast is nested in the combination forecasts. All other pairs are non-nested.

Table 10
Modified Diebold–Mariano tests and McCracken F-test.

	AR(I)MA		AR(I)MA-GARCH		ARFIMA		RS	
	DM test	MC-F test	DM test	MC-F test	DM test	MC-F test	DM test	MC-F test
AR(I)MA					2.216**		1.181	
AR(I)MA–GARCH		−21.214			0.588		0.334	
ARFIMA	−2.216		−0.588				−0.266	
RS	−1.181		−0.334		0.266			
HAR	−0.896		−0.166		0.529		0.186	
ED	−0.663		−0.178		0.245		0.075	
Equal		31.383***		53.053***		82.842***		73.027
Schwarz		31.537***		53.209***		82.999***		73.184
Time-varying		−21.876*		−0.667***		28.488***		18.882

	HAR		ED		Equal		Schwarz		Time-varying	
	DM test	MC-F test	DM test	MC-F test	DM test	MC-F test	DM test	MC-F test	DM test	MC-F test
AR(I)MA	0.896		0.663							
AR(I)MA-GARCH	0.166		0.178							
ARFIMA	−0.529		−0.245							
RS	−0.186		−0.075							
HAR			0.112							
ED	−0.112								1.848***	
Equal		63.909***		68.601***			−0.177		1.897**	
Schwarz		64.065***		68.758***	0.177					
Time-varying		9.958***		14.550***	−1.848		−1.897			

The table presents the McCracken F-statistics and the t-statistics for the modified Diebold–Mariano test for nested and non-nested models, respectively. The McCracken F-statistics are reported only if the column model is nested in the row model. The null hypothesis that the models in the rows perform just as well as the models in the columns is tested against a more accurate alternative model in the rows. The models considered are the AR(I)MA, AR(I)MA-GARCH, ARFIMA, regime-switching (RS), HAR, economic determinants (ED) and random walk models, in addition to the combination forecasts obtained using Schwarz, equal, and time-varying weights.
* Indicate rejection of the null hypothesis at the 10% significance level.
** Indicate rejection of the null hypothesis at the 5% significance level.
*** Indicate rejection of the null hypothesis at the 1% significance level.

ARFIMA, 4 for regime-switching, 5 for HAR, 6 for the economic determinants model, 7 for the Schwarz-weighted combination forecast, 8 for the equal combination forecast, and 9 for the time-varying combination forecast. The null hypothesis of an equal forecasting accuracy is tested against the alternative that the forecasting model i performs better than the benchmark model j, i.e., $E(d_{ijt}) > 0$. We compute the loss function in terms of both the mean square error of the forecast and the mean absolute error. The Diebold–Mariano test statistic is defined as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \sim N(0, 1), \tag{19}$$

where $\bar{d}$ is the sample average of the loss differential and $\hat{f}_d(0)$ is the estimate of the spectral density at frequency zero. For h-step-ahead forecasts, the modified Diebold–Mariano test statistic corrects for small sample sizes and the autocorrelation of the loss differential following a Student-t distribution with $T-1$ degrees of freedom, and is equal to

$$DM_{\text{mod}} = \left[\frac{T+1-2h+h(h-1)/T}{T}\right] DM. \tag{20}$$

Table 10 reports the t-statistics for the null hypothesis. In terms of mean squared errors, the model in row i performs just as well as that in column j.[9] The null hypothesis of equal errors is accepted for the majority of the pairwise comparisons. Interestingly, though, at the 95% confidence level, the equal- and Schwarz-weighted combination forecasts perform better than all but the AR(I)MA-GARCH model, and the AR(I)MA model also outperforms the ARFIMA model.

[9] The test results are similar when the loss function of the modified Diebold–Mariano test is defined with regard to the mean absolute error.

4.2. Economic significance

From a practical perspective, the presence of predictability in the evolution of the series enables market participants to form profitable trading strategies, thus constituting a direct test of the efficient market hypothesis. In addition to the statistical evaluation measures, we also investigate whether the predictability of the MFIC is significant enough to generate abnormal profits. Following Guo (2000), Goncalves and Guidolin (2006) and Harvey and Whaley (1992), we evaluate the out-of-sample forecasting performances of different forecasting models based on the profitability of a trading strategy. Although, in contrast to other relevant studies (e.g., Driessen, Maenhout, & Vilkov, 2009), our trading strategy is based on the idea of 'dispersion trading', we focus on exploiting daily changes in the MFIC (i.e., daily changes in the implied volatilities of stock and index options), rather than on differences between the implied and realized volatilities of stock and index options. A long (short) dispersion trade involves short (long) positions near ATM straddles on the S&P 100 index and long (short) positions on a portfolio near ATM straddles on S&P 100 component stocks. Straddles involve buying or selling equal amounts of call and put options with the same maturities and strike prices, and provide an effective way of trading based on changes in the implied volatility (see Brooks & Oozeer, 2002; Guo, 2000; and Ni, Pan, & Poteshman, 2008). A long (short) dispersion trade will generate profits if the MFIC decreases (increases), i.e., if the change in the implied volatilities of the stock options is larger (smaller) than the change in the index options' implied volatility. Thus, an investor may accumulate profits by betting on directional changes in the MFIC.

Our correlation trading strategy is as follows. If the MFIC is expected to increase (decrease) from times t to $t+1$, the investor takes a short (long) position on the dispersion trade. To build the dispersion trade portfolios, on each day for each asset and the index, we choose the call and put options of the shortest maturity with at least one call and one put with the same strike price. If more than one pair satisfies the criterion, we choose the one with the moneyness closest to unity. We discard options with maturities of less than seven days, ask prices that are lower than the bid prices, non-positive bid prices, moneyness levels higher than 1.15 or lower than 0.85, and zero open interest. We impose a restriction that \$1000 worth of options must be bought and sold each day, with the position being liquidated the next day. The funds may be invested freely in the riskless interest rate, but the profits are not reinvested the next day. The trading exercise is repeated for every day of the out-of-sample period.

The value of a portfolio unit on day t is computed as follows:

$$V_t = \begin{cases} -\left(C_{I,t}+P_{I,t}\right)+\sum_{i=1}^{N} n_{i,t}\left(C_{i,t}+P_{i,t}\right) \\ \quad \text{if long dispersion trade} \\ \left(C_{I,t}+P_{I,t}\right)-\sum_{i=1}^{N} n_{i,t}\left(C_{i,t}+P_{i,t}\right) \\ \quad \text{if short dispersion trade,} \end{cases} \tag{21}$$

where $C_{i,t}$ is the call price on stock i, $P_{i,t}$ is the put price on stock i, $C_{I,t}$ is the call price on the index, $P_{I,t}$ is the put price on the index, $n_{i,t} = \frac{N_{i,t}S_{I,t}}{\sum_{i=1}^{N} N_{i,t}S_{i,t}}$, $S_{i,t}$ is the closing price of stock i, $S_{I,t}$ is the closing price of the index, and $N_{i,t}$ is the number of shares outstanding in stock i for all variables on day t. We assume that \$1000 is invested in the portfolio each day, i.e., $X_t = 1000/|V_t|$ portfolio units are bought/sold. If the portfolio requires funds for its initiation (i.e., $V_t > 0$), the net gain of the portfolio is:

$$NG_{t+1} = X_t(V_{t+1} - V_t). \tag{22}$$

If the portfolio generates inflows at its initiation (i.e., $V_t < 0$), the net gain of the portfolio is:

$$NG_{t+1} = X_t\left(V_{t+1} - V_t\right) + 2000\left(e^{r_t/252} - 1\right). \tag{23}$$

To avoid noisy signals, the trading strategy is initiated only if the forecasted change in the MFIC is above the predefined filter levels of 0.1%, 0.5% and 1%.

Table 11 depicts the out-of-sample average returns, annualized Sharpe's ratios and Leland's alpha values. Whereas Sharpe's ratio is an appropriate measure of the profitability in the case of normal returns, Leland's (1999) alpha allows for non-normal trading strategy returns by taking into account higher-order moments of the return distribution. In accordance with the asset-pricing model of Rubinstein (1976), it is assumed that the returns on the market portfolio are i.i.d. over time and that the power utility of the agent is characterized by a constant risk aversion coefficient γ. Under these assumptions, a marginal utility-adjusted abnormal return measure for the trading strategy, defined as Leland's (1999) alpha, is derived as follows:

$$A = E\left[\frac{G_{t+1}}{1000}\right] - r_t - B\left(E\left[r_{m,t}\right] - r_t\right), \tag{24}$$

where $r_{m,t}$ is the return on the market portfolio, r_t is the risk-free rate, $B = \frac{\mathrm{Cov}\left(E\left[G_{t+1}/1000\right], -\left(1+r_{m,t}\right)^{-\gamma}\right)}{\mathrm{Cov}\left(r_{m,t}, -\left(1+r_{m,t}\right)^{-\gamma}\right)}$ is the market price of risk, and $\gamma = \frac{\ln\left(E\left[1+r_{m,t}\right]\right)-\ln(1+r_t)}{\mathrm{Var}\left(\ln\left(1+r_{m,t}\right)\right)}$ is a measure of the risk aversion. The daily USD Libor and S&P 100 returns are used as the risk-free rate and the market returns, respectively. Finally, because the trading strategy returns are found to be stationary and non-normal (based on the unreported results), the statistical significance of our results is assessed using bootstrapped 95% confidence intervals based on the method of Politis and Romano (1994). The average block size is set to ten, and 1000 bootstrap repetitions are used.

The reported results with no filter indicate that the trading strategies based on the AR(I)MA-GARCH model and the equal- and Schwarz-weighted combination forecasts produce significant positive average returns over the out-of-sample period. The highest return, Sharpe's ratio and Leland's alpha are obtained using the AR(I)MA-GARCH model. Moreover, when filters are applied, these three models yield significant positive returns across all filters.

The above results call into question the efficiency of the S&P100 index and the stock option market. However, the incorporation of transaction costs allows the profitability of the trading strategy and the market efficiency hypothesis to be assessed thoroughly. Specifically, we approximate the transaction costs by assuming that an investor buys options at the ask price and sells at the bid price. Table 12 reports the out-of-sample average returns, annualized Sharpe's ratios and Leland's alpha values, along with the corresponding bootstrapped confidence intervals for the correlation trading strategy based on alternative forecasting models and filters after incorporating transaction costs. Unsurprisingly, transaction costs have a major impact on the profitability of the trading strategies. The average daily return, Sharpe's ratio and Leland's alpha are all significantly negative across all forecasting models and filters applied. Our results extend the findings of Bernales and Guidolin (2014), Driessen et al. (2009), Goyal and Saretto (2009) and Neumann and Skiadopoulos (2013), who also note the significant economic impact of bid–ask spreads in option trading.

Most importantly, our results imply that only investors who face zero or limited transaction costs (e.g., market makers) can generate abnormal profits through the correct prediction of changes in the MFIC. The elimination of profitability when transaction costs are taken into account

Table 11
Trading strategy based on out-of-sample forecasts, without transaction costs.

Panel A						
Model	No filter			Filter 0.1%		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	0.195%	0.425	0.002	0.192%	0.417	0.002
C.I. (95%)	*(−0.008%, 0.401%)*	*(−0.092, 0.918)*	*(0.000, 0.004)*	*(−0.021%, 0.407%)*	*(−0.073, 0.911)*	*(0.000, 0.004)*
AR(I)MA-GARCH	0.335%*	0.746*	0.003*	0.335%*	0.746*	0.003*
C.I. (95%)	*(0.108%, 0.582%)*	*(0.229, 1.197)*	*(0.001, 0.006)*	*(0.109%, 0.560%)*	*(0.274, 1.218)*	*(0.001, 0.006)*
ARFIMA	0.188%	0.408	0.002	0.186%	0.405	0.002
C.I. (95%)	*(−0.037%, 0.399%)*	*(−0.111, 0.973)*	*(0.000, 0.004)*	*(−0.037%, 0.408%)*	*(−0.094, 0.977)*	*(−0.001, 0.004)*
RS	0.201%	0.438	0.002	0.193%	0.420	0.002
C.I. (95%)	*(−0.055%, 0.441%)*	*(−0.097, 1.037)*	*(−0.001, 0.004)*	*(−0.042%, 0.433%)*	*(−0.099, 0.997)*	*(−0.001, 0.004)*
ED	0.181%	0.393	0.002	0.177%	0.382	0.002
C.I. (95%)	*(−0.054%, 0.398%)*	*(−0.107, 0.955)*	*(0.000, 0.004)*	*(−0.047%, 0.378%)*	*(−0.084, 0.966)*	*(0.000, 0.004)*
HAR	−0.009%	*−0.042*	*0.000*	−0.008%	*−0.040*	*0.000*
C.I. (95%)	*(−0.240%, 0.236%)*	*(−0.553, 0.531)*	*(−0.002, 0.002)*	*(−0.248%, 0.222%)*	*(−0.562, 0.489)*	*(−0.003, 0.002)*
Combination forecasts						
Schwarz	0.271%*	0.600*	0.003*	0.273%*	0.604*	0.003*
C.I. (95%)	*(0.061%, 0.471%)*	*(0.119, 1.105)*	*(0.001, 0.005)*	*(0.053%, 0.470%)*	*(0.139, 1.095)*	*(0.001, 0.005)*
Equal	0.268%*	0.593*	0.003*	0.272%*	0.602*	0.003*
C.I. (95%)	*(0.061%, 0.490%)*	*(0.116, 1.090)*	*(0.001, 0.005)*	*(0.060%, 0.484%)*	*(0.109, 1.086)*	*(0.001, 0.005)*
Time-varying	0.060%	0.115	0.000	0.058%	0.110	0.000
C.I. (95%)	*(−0.157%, 0.292%)*	*(−0.348, 0.684)*	*(−0.002, 0.003)*	*(−0.159%, 0.275%)*	*(−0.419, 0.631)*	*(−0.002, 0.003)*

Panel B						
Model	Filter 0.5%			Filter 1%		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	0.191%	0.417	0.002	0.188%	0.409	0.002
C.I. (95%)	*(−0.029%, 0.394%)*	*(−0.051, 0.887)*	*(0.000, 0.004)*	*(−0.024%, 0.382%)*	*(−0.087, 0.898)*	*(0.000, 0.004)*
AR(I)MA-GARCH	0.329%*	0.735*	0.003*	0.336%*	0.755*	0.003*
C.I. (95%)	*(0.105%, 0.576%)*	*(0.204, 1.218)*	*(0.001, 0.006)*	*(0.100%, 0.573%)*	*(0.263, 1.230)*	*(0.001, 0.006)*
ARFIMA	0.211%	0.463	0.002	0.219%	0.485	0.002
C.I. (95%)	*(−0.009%, 0.434%)*	*(−0.038, 1.054)*	*(0.000, 0.004)*	*(−0.005%, 0.439%)*	*(−0.033, 1.112)*	*(0.000, 0.004)*
RS	0.194%	0.424	0.002	0.154%	0.336	0.001
C.I. (95%)	*(−0.047%, 0.425%)*	*(−0.143, 1.003)*	*(−0.001, 0.004)*	*(−0.096%, 0.374%)*	*(−0.206, 0.958)*	*(−0.001, 0.004)*
ED	0.161%	0.347	0.002	0.179%	0.390	0.002
C.I. (95%)	*(−0.057%, 0.364%)*	*(−0.177, 0.857)*	*(−0.001, 0.003)*	*(−0.037%, 0.398%)*	*(−0.142, 0.951)*	*(−0.001, 0.004)*
HAR	−0.042%	*−0.121*	*−0.001*	−0.042%	*−0.122*	*−0.001*
C.I. (95%)	*(−0.286%, 0.200%)*	*(−0.695, 0.456)*	*(−0.003, 0.002)*	*(−0.298%, 0.197%)*	*(−0.642, 0.460)*	*(−0.003, 0.002)*
Combination forecasts						
Schwarz	0.284%*	0.631*	0.003*	0.265%*	0.591*	0.002*
C.I. (95%)	*(0.073%, 0.502%)*	*(0.144, 1.119)*	*(0.001, 0.005)*	*(0.036%, 0.469%)*	*(0.072, 1.096)*	*(0.000, 0.005)*
Equal	0.276%*	0.613*	0.003*	0.265%*	0.590*	0.002*
C.I. (95%)	*(0.076%, 0.499%)*	*(0.143, 1.131)*	*(0.000, 0.005)*	*(0.055%, 0.480%)*	*(0.109, 1.093)*	*(0.000, 0.005)*
Time-varying	0.056%	0.106	0.000	0.058%	0.111	0.000
C.I. (95%)	*(−0.177%, 0.283%)*	*(−0.416, 0.633)*	*(−0.002, 0.003)*	*(−0.171%, 0.278%)*	*(−0.378, 0.632)*	*(−0.002, 0.003)*

The table reports the average daily return, the annualized Sharpe's ratio, and Leland's alpha, corresponding to the MFIC forecasts from the models and to the combination forecasts obtained using the Schwarz, equal and time-varying weights. The results are reported for the trading strategy without filters and with 0.1% (Panel A), 0.5% and 1% (Panel B) filters. The bootstrapped 95% confidence intervals (CI) are reported in parentheses.

* Indicates rejection of the null hypothesis of a zero average return, Sharpe's ratio or Leland's alpha, at a 5% significance level.

can be attributed to the simultaneous trading position of the investor for all constituent stocks of the index. An alternative strategy that significantly reduces transaction costs and enhances the profitability could be to replicate the dispersion trade partially, by holding long/short positions on options only on certain representative stocks of the index (e.g., those representing 75% of the index market capitalization).

5. Robustness checks

Because our dataset covers an extensive period, from 1996 to 2010, an assessment of the robustness of our results across different sample periods is indispensable. To this end, we use expanding in-sample windows of three, six, nine and twelve years to produce out-of-sample forecasts for the following three years. Thus, this method creates a series of non-overlapping forecasted values. Specifically, the in-sample periods are: from January 1996 to December 1998, from January 1996 to December 2001, from January 1996 to December 2004, and from January 1996 to December 2007, corresponding to the following out-of-sample forecasting periods: from January 1999 to December 2001, from January 2002 to December 2004, from January 2005 to December 2007, and from January 2008 to October 2010, respectively.

Table 12
Trading strategy based on out-of-sample forecasts, with transaction costs.

Panel A						
Model	No filter			Filter 0.1%		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	−12.531%*	−8.530*	−0.125*	−12.528%*	−8.524*	−0.125*
C.I. (95%)	(*−13.910%, −11.288%*)	(*−10.987, −7.007*)	(*−0.138, −0.114*)	(*−13.870%, −11.322%*)	(*−10.973, −6.931*)	(*−0.139, −0.113*)
AR(I)MA-GARCH	−12.234%*	−9.889*	−0.123*	−12.253%*	−9.889*	−0.122*
C.I. (95%)	(*−13.456%, −11.223%*)	(*−11.095, −8.852*)	(*−0.134, −0.112*)	(*−13.344%, −11.232%*)	(*−11.184, −8.870*)	(*−0.133, −0.112*)
ARFIMA	−13.362%*	−3.692*	−0.134*	−13.311%*	−3.672*	−0.133*
C.I. (95%)	(*−16.158%, −11.218%*)	(*−10.836, −2.716*)	(*−0.162, −0.112*)	(*−16.163%, −11.230%*)	(*−10.801, −2.680*)	(*−0.163, −0.112*)
RS	−13.684%*	−3.774*	−0.137*	−13.693%*	−3.767*	−0.137*
C.I. (95%)	(*−16.475%, −11.546%*)	(*−10.804, −2.757*)	(*−0.167, −0.116*)	(*−16.600%, −11.484%*)	(*−10.807, −2.727*)	(*−0.165, −0.115*)
ED	−12.930%*	−7.700*	−0.129*	−12.903%*	−7.683*	−0.129*
C.I. (95%)	(*−14.409%, −11.571%*)	(*−10.869, −6.046*)	(*−0.146, −0.116*)	(*−14.408%, −11.604%*)	(*−10.821, −6.146*)	(*−0.144, −0.115*)
HAR	−13.178%*	−7.701*	−0.132*	−13.128%*	−7.664*	−0.131*
C.I. (95%)	(*−14.898%, −11.786%*)	(*−10.671, −6.130*)	(*−0.148, −0.117*)	(*−14.770%, −11.736%*)	(*−10.581, −6.219*)	(*−0.149, −0.116*)
Combination forecasts						
Schwarz	−12.585%*	−8.451*	−0.126*	−12.551%*	−8.426*	−0.125*
C.I. (95%)	(*−14.065%, −11.354%*)	(*−10.813, −6.930*)	(*−0.140, −0.114*)	(*−13.967%, −11.322%*)	(*−10.801, −6.856*)	(*−0.139, −0.114*)
Equal	−12.583%*	−8.451*	−0.126*	−12.569%*	−8.435*	−0.126*
C.I. (95%)	(*−14.028%, −11.392%*)	(*−10.851, −6.885*)	(*−0.138, −0.114*)	(*−13.896%, −11.348%*)	(*−10.808, −6.954*)	(*−0.141, −0.113*)
Time-varying	−12.966%*	−8.326*	−0.130*	−12.954%*	−8.308*	−0.129*
C.I. (95%)	(*−14.518%, −11.709%*)	(*−10.683, −6.867*)	(*−0.145, −0.117*)	(*−14.311%, −11.690%*)	(*−10.675, −6.934*)	(*−0.144, −0.117*)

Panel B						
Model	Filter 0.5%			Filter 1%		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	−12.412%*	−8.470*	−0.124*	−12.336%*	−8.418*	−0.123*
C.I. (95%)	(*−13.863%, −11.167%*)	(*−10.953, −6.755*)	(*−0.138, −0.113*)	(*−13.694%, −11.171%*)	(*−10.844, −6.913*)	(*−0.136, −0.111*)
AR(I)MA-GARCH	−12.141%*	−9.826*	−0.122*	−11.996%*	−9.726*	−0.120*
C.I. (95%)	(*−13.238%, −11.055%*)	(*−11.035, −8.835*)	(*−0.133, −0.111*)	(*−13.131%, −10.892%*)	(*−10.972, −8.706*)	(*−0.132, −0.109*)
ARFIMA	−13.034%*	−3.607*	−0.131*	−12.823%*	−3.551*	−0.128*
C.I. (95%)	(*−15.992%, −11.107%*)	(*−10.754, −2.690*)	(*−0.158, −0.110*)	(*−15.663%, −10.749%*)	(*−10.617, −2.637*)	(*−0.156, −0.107*)
RS	−13.521%*	−3.730*	−0.136*	−13.400%*	−3.697*	−0.134*
C.I. (95%)	(*−16.500%, −11.446%*)	(*−10.623, −2.774*)	(*−0.165, −0.113*)	(*−16.511%, −11.220%*)	(*−10.587, −2.722*)	(*−0.163, −0.114*)
ED	−12.819%*	−7.634*	−0.128*	−12.645%*	−7.541*	−0.126*
C.I. (95%)	(*−14.428%, −11.490%*)	(*−10.761, −6.128*)	(*−0.144, −0.115*)	(*−14.183%, −11.274%*)	(*−10.615, −6.058*)	(*−0.142, −0.112*)
HAR	−13.022%*	−7.616*	−0.130*	−12.859%*	−7.525*	−0.129*
C.I. (95%)	(*−14.785%, −11.650%*)	(*−10.600, −6.165*)	(*−0.147, −0.115*)	(*−14.478%, −11.407%*)	(*−10.421, −5.987*)	(*−0.146, −0.115*)
Combination forecasts						
Schwarz	−12.373%*	−8.365*	−0.124*	−12.218%*	−8.285*	−0.122*
C.I. (95%)	(*−13.632%, −11.171%*)	(*−10.827, −6.859*)	(*−0.138, −0.112*)	(*−13.635%, −11.087%*)	(*−10.702, −6.691*)	(*−0.135, −0.111*)
Equal	−12.369%*	−8.362*	−0.124*	−12.210%*	−8.280*	−0.122*
C.I. (95%)	(*−13.755%, −11.168%*)	(*−10.810, −6.873*)	(*−0.137, −0.112*)	(*−13.475%, −11.070%*)	(*−10.717, −6.807*)	(*−0.136, −0.111*)
Time-varying	−12.740%*	−8.213*	−0.128*	−12.530%*	−8.116*	−0.125*
C.I. (95%)	(*−14.048%, −11.498%*)	(*−10.544, −6.860*)	(*−0.142, −0.115*)	(*−13.928%, −11.325%*)	(*−10.512, −6.723*)	(*−0.139, −0.113*)

The table reports the average daily return, the annualized Sharpe's ratio, and Leland's alpha corresponding to the MFIC forecasts from the models and to the combination forecasts obtained using the Schwarz, equal and time-varying weights, after accounting for transaction costs. The results are reported for the trading strategy without filters and with 0.1% (Panel A), 0.5% and 1% filters (Panel B). The bootstrapped 95% confidence intervals (CI) are reported in parentheses.

* Indicates rejection of the null hypothesis of a zero average return, Sharpe's ratio or Leland's alpha at a 5% significance level.

5.1. Evaluation of the out-of-sample performances

Table 13 shows the forecasting performances of the models over the different sampling periods, together with the average values over the out-of-sample period from January 1, 1999, to October 29, 2010. We observe that, although the RMSE and the MAE do not vary significantly across models in the same period, they do vary across the different sampling schemes. Specifically, the RMSE is lower during the first two forecasting periods than in the main estimation window; however, the last forecasting period, which includes the 2008 crisis, obtains the largest values in terms of both the RMSE and the MAE. With regard to directional accuracy, the results do not vary significantly among the different sample windows, including the main estimation window. Across all periods and models, with the exception of the AR(I)MA-GARCH model for the last forecasting period, which includes the 2008 crisis, the PT test suggests that the direction of the change of the MFIC series is indeed predictable.

The last column of Table 13 presents the average values of the statistical measures for evaluating the performances of the competing models for the out-of-sample period 1999–2010, where the forecasted values are obtained as described above. The time-varying model and the Schwartz combination forecast result in the minimum RMSE and MAE values, respectively. In terms of the best performing model on average, the weighted combination forecast produces superior forecasts across the various sample periods, whereas the Schwarz- and equal-weighted combination forecasts were the best performing models for the main estimation window. Our results are in line with a large body of literature that suggests that the

Table 13
Robustness checks: evaluation of out-of-sample performances for different sub-samples.

	1999–2001					2002–2004				
	RMSE	MAE	MSE-F	MCP	PT test	RMSE	MAE	MSE-F	MCP	PT test
AR(I)MA	14.618%	11.191%	73.006***	60.347%	5.620***	12.462%	9.515%	99.512***	61.772%	6.396***
AR(I)MA-GARCH	14.536%	11.096%	82.313***	61.148%	6.247***	12.423%	9.470%	104.823***	62.566%	6.828***
ARFIMA	15.290%	11.743%	2.264*	54.206%	2.312**	13.040%	10.022%	25.937***	54.630%	2.532**
RS	14.629%	11.170%	71.700***	58.611%	4.889***	12.600%	9.581%	80.956***	56.878%	6.382***
HAR	15.087%	11.601%	22.662***	56.208%	3.434***	12.998%	10.004%	30.907***	56.349%	3.517***
ED	15.040%	11.459%	27.486***	59.680%	5.296***	18.206%	12.284%	−351.456	60.847%	5.869***
Equal	14.560%	11.107%	79.535***	60.881%	5.933***	12.458%	9.508%	100.043***	60.317%	5.801***
Time-varying	14.620%	11.139%	72.785***	60.214%	5.989***	12.447%	9.532%	101.578***	61.376%	6.255***
Schwarz	14.552%	11.101%	80.379***	60.881%	6.006***	12.455%	9.506%	100.455***	60.582%	5.944***
Random walk	15.313%	11.745%				13.264%	10.215%			

	2005–2007					2008–2010				
	RMSE	MAE	MSE-F	MCP	PT test	RMSE	MAE	MSE-F	MCP	PT test
AR(I)MA	18.189%	12.212%	73.537***	61.406%	6.064***	22.978%	14.685%	44.520***	54.278%	2.375**
AR(I)MA-GARCH	18.251%	12.245%	67.955***	60.743%	5.642***	23.336%	14.933%	20.344***	53.156%	1.586
ARFIMA	18.271%	12.461%	66.207***	59.284%	5.097***	22.784%	14.506%	58.077***	56.522%	3.428***
RS	19.246%	12.694%	−14.292	58.090%	6.390***	24.357%	15.197%	−42.817	55.259%	2.203**
HAR	18.294%	12.395%	64.121***	60.477%	5.841***	22.665%	14.491%	66.604***	56.522%	3.653***
ED	18.206%	12.284%	71.993***	60.477%	5.682***	23.138%	14.778%	33.587***	55.680%	2.973***
Equal	18.163%	12.183%	75.880***	61.671%	6.754***	22.989%	14.589%	43.783***	55.259%	2.782***
Time-varying	18.156%	12.202%	76.561***	60.875%	5.864***	22.864%	14.517%	52.438***	56.381%	3.779***
Schwarz	18.168%	12.186%	75.431***	61.538%	6.676***	22.998%	14.595%	43.120***	55.400%	3.003***
Random walk	19.061%	12.916%				23.651%	15.043%			

	Average				
	RMSE	MAE	MSE-F	MCP	PT test
AR(I)MA	17.434%	11.862%	64.727***	59.5222%	10.527***
AR(I)MA-GARCH	17.540%	11.894%	54.927***	59.4886%	10.754***
ARFIMA	17.647%	12.150%	45.263***	56.1575%	6.124***
RS	18.183%	12.118%	−0.909*	57.2342%	10.422***
HAR	17.564%	12.090%	52.776***	57.4024%	7.692***
ED	17.604%	12.002%	49.097***	59.2194%	9.895***
Equal	17.418%	11.809%	66.255***	59.5895%	10.799***
Time-varying	17.387%	11.810%	69.142***	59.7577%	10.827***
Schwarz	17.420%	11.809%	66.045***	59.6568%	10.918***
Random walk	18.172%	12.444%			

Panel A presents the root mean square error (RMSE) and the mean absolute error (MAE) in percentage terms, corresponding to the AR(I)MA, AR(I)MA-GARCH, ARFIMA, regime-switching (RS), HAR, economic determinants (ED) and random walk models, in addition to the combination forecasts obtained using equal, time-varying and Schwarz weights. The F-statistics of the McCracken test for the MSE are also reported. The null hypothesis that the forecasting accuracy of a given model is equal to that of the random walk model is tested against the alternative of a more accurate forecasting model. The mean correct prediction (MCP) represents the number of times that the actual change in the MFIC is predicted correctly by the forecasted series. The PT test examines the independence between the actual and forecasted values. The results are reported both across different forecast periods and on average.

* Indicate rejection of the null hypothesis at the 10% significance level.

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

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

forecasts from combination models are more accurate than those from the individual model specifications.

The statistics reported for the McCracken (2007) F-type test suggest that the null hypothesis is rejected strongly for all models except for the regime-switching model, in which the null hypothesis of an equal forecasting accuracy with the random walk is rejected only at the 10% significance level. In addition, the PT test also suggests that the models are able to track a predictable pattern in the directional changes of the MFIC series. Overall, our results across different forecasting periods suggest the existence of predictability in the MFIC series, which is consistent with the results obtained from the main estimation window, and thus enhances the rejection of the market efficiency hypothesis.

5.2. Economic significance

Table 14 presents the results of the correlation trading strategy over the four out-of-sample periods without accounting for transaction costs. For the sake of brevity, we report only the results without filters. During all but the last forecast period, at least one model produces significantly positive returns, although these models differ across the various sub-periods. Interestingly, the majority of the models obtain significant positive returns when forecasting the period before the 2008 crisis, whereas none of them obtain significant returns during the last forecast period, which includes the 2008 crisis. The last column of Table 14 presents the average return, Sharpe's ratio and Leland's alpha across the four out-of-sample periods. The highest return is obtained by the AR(I)MA-GARCH

Table 14
Robustness checks: trading strategy based on out-of-sample forecasts, without transaction costs.

Model	1999–2001			2002–2004		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	−0.191%	−0.797	−0.002	−0.009%	−0.044	0.000
C.I. (95%)	*(−0.472%, 0.101%)*	*(−1.951, 0.256)*	*(−0.005, 0.001)*	*(−0.310%, 0.307%)*	*(−1.014, 0.869)*	*(−0.004, 0.003)*
AR(I)MA-GARCH	0.162%	0.532	0.001	0.172%	0.506	0.002
C.I. (95%)	*(−0.097%, 0.428%)*	*(−0.415, 1.525)*	*(−0.001, 0.004)*	*(−0.118%, 0.471%)*	*(−0.395, 1.418)*	*(−0.001, 0.005)*
ARFIMA	0.277%	0.965	0.003	0.563%*	1.705*	0.006*
C.I. (95%)	*(−0.031%, 0.585%)*	*(−0.168, 2.088)*	*(0.000, 0.006)*	*(0.231%, 0.928%)*	*(0.632, 2.673)*	*(0.002, 0.009)*
RS	−0.137%	−0.591	−0.002	−0.055%	−0.184	−0.001
C.I. (95%)	*(−0.400%, 0.118%)*	*(−1.610, 0.379)*	*(−0.004, 0.001)*	*(−0.333%, 0.238%)*	*(−0.954, 0.728)*	*(−0.003, 0.002)*
ED	0.439%*	1.580*	0.004*	0.122%	0.355	0.001
C.I. (95%)	*(0.140%, 0.759%)*	*(0.442, 2.687)*	*(0.001, 0.007)*	*(−0.173%, 0.419%)*	*(−0.522, 1.325)*	*(−0.002, 0.004)*
HAR	−0.197%	*−0.819*	*−0.002*	−0.016%	*−0.067*	*0.000*
C.I. (95%)	*(−0.467%, 0.041%)*	*(−1.718, 0.145)*	*(−0.005, 0.000)*	*(−0.290%, 0.249%)*	*(−0.892, 0.778)*	*(−0.003, 0.003)*
Combination forecasts						
Schwarz	−0.064%	−0.320	−0.001	0.144%	0.419	0.001
C.I. (95%)	*(−0.325%, 0.201%)*	*(−1.379, 0.720)*	*(−0.003, 0.002)*	*(−0.119%, 0.412%)*	*(−0.420, 1.301)*	*(−0.001, 0.004)*
Equal	−0.065%	−0.322	−0.001	0.148%	0.432	0.001
C.I. (95%)	*(−0.343%, 0.211%)*	*(−1.353, 0.621)*	*(−0.004, 0.002)*	*(−0.109%, 0.422%)*	*(−0.392, 1.267)*	*(−0.001, 0.004)*
Time-varying	0.134%	0.426	0.001	−0.044%	−0.152	0.000
C.I. (95%)	*(−0.136%, 0.381%)*	*(−0.491, 1.370)*	*(−0.001, 0.004)*	*(−0.335%, 0.243%)*	*(−1.140, 0.694)*	*(−0.003, 0.002)*
Model	**2005–2007**			**2008–2010**		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	0.493%*	1.394*	0.005*	−0.048%	−0.086	−0.001
C.I. (95%)	*(0.175%, 0.840%)*	*(0.639, 2.072)*	*(0.002, 0.009)*	*(−0.641%, 0.532%)*	*(−0.997, 0.888)*	*(−0.006, 0.005)*
AR(I)MA-GARCH	0.585%*	1.666*	0.006*	0.318%	0.525	0.003
C.I. (95%)	*(0.261%, 0.951%)*	*(0.932, 2.295)*	*(0.003, 0.009)*	*(−0.211%, 0.910%)*	*(−0.459, 1.340)*	*(−0.002, 0.009)*
ARFIMA	0.204%	0.545	0.002	−0.051%	−0.092	−0.001
C.I. (95%)	*(−0.097%, 0.487%)*	*(−0.198, 1.761)*	*(−0.001, 0.005)*	*(−0.601%, 0.511%)*	*(−1.001, 0.907)*	*(−0.006, 0.005)*
RS	0.436%*	1.226*	0.004*	−0.094%	−0.163	−0.001
C.I. (95%)	*(0.106%, 0.836%)*	*(0.259, 1.961)*	*(0.001, 0.008)*	*(−0.673%, 0.511%)*	*(−1.045, 0.893)*	*(−0.007, 0.004)*
ED	0.218%	0.584	0.002	−0.310%	−0.524	−0.003
C.I. (95%)	*(−0.057%, 0.493%)*	*(−0.172, 1.702)*	*(−0.001, 0.005)*	*(−0.876%, 0.284%)*	*(−1.415, 0.406)*	*(−0.009, 0.003)*
HAR	0.372%*	*1.037**	*0.004**	−0.018%	*−0.036*	*0.000*
C.I. (95%)	*(0.048%, 0.715%)*	*(0.158, 1.789)*	*(0.001, 0.007)*	*(−0.620%, 0.587%)*	*(−0.955, 1.058)*	*(−0.007, 0.005)*
Combination forecasts						
Schwarz	0.500%*	1.415*	0.005*	0.030%	0.044	0.000
C.I. (95%)	*(0.195%, 0.873%)*	*(0.619, 2.128)*	*(0.002, 0.009)*	*(−0.525%, 0.593%)*	*(−0.867, 0.996)*	*(−0.006, 0.006)*
Equal	0.514%*	1.458*	0.005*	0.020%	0.026	0.000
C.I. (95%)	*(0.188%, 0.910%)*	*(0.659, 2.144)*	*(0.002, 0.009)*	*(−0.534%, 0.566%)*	*(−0.812, 1.120)*	*(−0.006, 0.006)*
Time-varying	0.483%*	1.364*	0.005*	−0.158%	−0.271	−0.002
C.I. (95%)	*(0.159%, 0.843%)*	*(0.514, 2.067)*	*(0.001, 0.008)*	*(−0.751%, 0.440%)*	*(−1.187, 0.774)*	*(−0.008, 0.004)*

Model	Average		
	Average daily return	Annualized Sharpe ratio	Leland's alpha
AR(I)MA	0.061%	0.117	0.000
AR(I)MA-GARCH	0.309%	0.807	0.003
ARFIMA	0.248%	0.781	0.002
RS	0.038%	0.072	0.000
ED	0.117%	0.499	0.001
HAR	0.035%	0.029	0.000
Combination forecasts			
Schwarz	0.152%	0.390	0.001
Equal	0.154%	0.399	0.001
Time-varying	0.104%	0.342	0.001

The table reports the average daily return, the annualized Sharpe's ratio, and Leland's alpha corresponding to the MFIC forecasts from the models and to the combination forecasts obtained using the Schwarz, equal and time-varying weights. The results are reported for the trading strategy without filters. Bootstrapped 95% confidence intervals (CI) are reported in parentheses. The results are reported across different forecast periods and on average.

* Indicates the rejection of the null hypothesis of a zero average return, Sharpe's ratio or Leland's alpha at the 5% significance level.

specification, which is consistent with the results from the main estimation window. The models with the second, third and fourth highest average returns are the ARFIMA model and the equal- and Schwarz-weighted combination forecasts, respectively. Once transaction costs are taken into account, all of the models produce significant negative returns over the four out-of-sample periods (Table 15), similarly to the results for the main estimation window in Table 12. With the exception of the after-crisis forecast period, we find our main economic result – the existence

Table 15
Robustness checks: trading strategy based on out-of-sample forecasts, with transaction costs.

Model	1999–2001			2002–2004		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	−10.398%*	−12.336*	−0.104*	−14.339%*	−12.375*	−0.143*
C.I. (95%)	(−11.562%, −9.272%)	(−12.904, −11.847)	(−0.117, −0.093)	(−15.626%, −13.166%)	(−13.384, −11.523)	(−0.158, −0.132)
AR(I)MA-GARCH	−9.713%*	−12.434*	−0.097*	−13.872%*	−12.606*	−0.139*
C.I. (95%)	(−10.899%, −8.651%)	(−12.995, −11.944)	(−0.108, −0.087)	(−15.156%, −12.803%)	(−13.432, −11.885)	(−0.151, −0.128)
ARFIMA	−9.591%*	−12.144*	−0.096*	−13.123%*	−13.216*	−0.131*
C.I. (95%)	(−10.769%, −8.514%)	(−12.821, −11.662)	(−0.108, −0.085)	(−14.038%, −12.271%)	(−13.934, −12.537)	(−0.140, −0.123)
RS	−10.242%*	−12.454*	−0.102*	−14.341%*	−12.330*	−0.143*
C.I. (95%)	(−11.421%, −9.214%)	(−12.990, −12.006)	(−0.115, −0.093)	(−15.784%, −12.998%)	(−13.113, −11.669)	(−0.157, −0.131)
ED	−9.615%*	−12.327*	−0.096*	−14.320%*	−12.172*	−0.143*
C.I. (95%)	(−10.765%, −8.620%)	(−12.981, −11.794)	(−0.108, −0.086)	(−15.806%, −13.011%)	(−13.168, −11.412)	(−0.158, −0.130)
HAR	−10.298%*	−12.228*	−0.103*	−14.265%*	−12.435*	−0.143*
C.I. (95%)	(−11.456%, −9.059%)	(−12.849, −11.698)	(−0.116, −0.092)	(−15.623%, −12.960%)	(−13.219, −11.804)	(−0.156, −0.130)
Combination forecasts						
Schwarz	−10.192%*	−12.354*	−0.102*	−14.049%*	−12.560*	−0.141*
C.I. (95%)	(−11.505%, −9.114%)	(−12.952, −11.862)	(−0.114, −0.091)	(−15.237%, −12.921%)	(−13.348, −11.909)	(−0.153, −0.130)
Equal	−10.197%*	−12.362*	−0.102*	−14.041%*	−12.553*	−0.140*
C.I. (95%)	(−11.388%, −9.082%)	(−12.962, −11.912)	(−0.114, −0.092)	(−15.285%, −12.959%)	(−13.344, −11.887)	(−0.152, −0.129)
Time-varying	−9.787%*	−12.491*	−0.098*	−14.443%*	−12.474*	−0.144*
C.I. (95%)	(−10.847%, −8.765%)	(−13.020, −12.033)	(−0.109, −0.088)	(−15.905%, −13.230%)	(−13.312, −11.812)	(−0.158, −0.133)
Model	2005–2007			2008–2010		
	Average daily return	Sharpe ratio	Leland's alpha	Average daily return	Sharpe ratio	Leland's alpha
AR(I)MA	−8.853%*	−9.261*	−0.089*	−16.496%*	−7.335*	−0.165*
C.I. (95%)	(−10.461%, −7.182%)	(−10.452, −8.100)	(−0.106, −0.071)	(−20.562%, −13.378%)	(−11.552, −5.936)	(−0.205, −0.134)
AR(I)MA-GARCH	−8.597%*	−9.516*	−0.086*	−15.392%*	−9.827*	−0.154*
C.I. (95%)	(−10.205%, −6.953%)	(−10.702, −8.308)	(−0.104, −0.069)	(−18.257%, −12.941%)	(−11.398, −8.900)	(−0.181, −0.130)
ARFIMA	−9.042%*	−9.160*	−0.091*	−16.153%*	−7.340*	−0.162*
C.I. (95%)	(−10.958%, −7.341%)	(−10.409, −8.072)	(−0.109, −0.072)	(−19.813%, −13.359%)	(−11.279, −5.957)	(−0.196, −0.132)
RS	−9.104%*	−9.029*	−0.091*	−20.396%*	−3.096*	−0.205*
C.I. (95%)	(−10.840%, −7.226%)	(−10.259, −7.835)	(−0.110, −0.071)	(−30.058%, −13.554%)	(−10.451, −2.614)	(−0.298, −0.137)
ED	−9.110%*	−9.214*	−0.091*	−17.385%*	−6.584*	−0.174*
C.I. (95%)	(−10.919%, −7.178%)	(−10.445, −8.074)	(−0.109, −0.073)	(−22.255%, −13.677%)	(−11.208, −5.457)	(−0.221, −0.140)
HAR	−9.096%*	−9.017*	−0.091*	−19.937%*	−3.036*	−0.200*
C.I. (95%)	(−10.989%, −7.360%)	(−10.295, −7.833)	(−0.109, −0.072)	(−29.766%, −13.380%)	(−10.803, −2.563)	(−0.300, −0.136)
Combination forecasts						
Schwarz	−9.024%*	−9.121*	−0.090*	−16.280%*	−7.326*	−0.163*
C.I. (95%)	(−10.816%, −7.087%)	(−10.402, −8.023)	(−0.107, −0.073)	(−20.292%, −13.250%)	(−11.304, −6.039)	(−0.201, −0.131)
Equal	−9.014%*	−9.115*	−0.090*	−16.310%*	−7.336*	−0.163*
C.I. (95%)	(−10.769%, −7.323%)	(−10.345, −8.034)	(−0.108, −0.073)	(−20.000%, −13.168%)	(−11.176, −6.020)	(−0.204, −0.134)
Time-varying	−8.943%*	−9.237*	−0.090*	−20.174%*	−3.073*	−0.203*
C.I. (95%)	(−10.712%, −7.179%)	(−10.413, −8.069)	(−0.108, −0.071)	(−30.511%, −13.594%)	(−11.138, −2.591)	(−0.289, −0.137)

Model	Average		
	Average daily return	Annualized Sharpe ratio	Leland's alpha
AR(I)MA	−12.522%	−10.327	−0.125
AR(I)MA-GARCH	−11.894%	−11.096	−0.119
ARFIMA	−11.977%	−10.465	−0.120
RS	−13.521%	−9.227	−0.136
ED	−12.607%	−10.074	−0.126
HAR	−13.399%	−9.179	−0.134
Combination forecasts			
Schwarz	−12.386%	−10.340	−0.124
Equal	−12.390%	−10.341	−0.124
Time-varying	−13.337%	−9.319	−0.134

The table reports the average daily return, annualized Sharpe's ratio, and Leland's alpha corresponding to the MFIC forecasts from the models and to the combination forecasts obtained using the Schwarz, equal and time-varying weights, after accounting for transaction costs. The results are reported for the trading strategy without filters. Bootstrapped 95% confidence intervals (CI) are reported in parentheses. The results are reported across different forecast periods and on average.

* Indicates the rejection of the null hypothesis of a zero average return, Sharpe's ratio or Leland's alpha at the 5% significance level.

of profitable correlation trading strategies based on MFIC forecasts that fade out when transaction costs are taken into account – to be robust across different in-sample sizes and forecast periods.

6. Conclusions

Having an understanding of the dynamics that govern the evolution of correlations is of vital importance for asset pricing theory and other financial applications. In this study, we examine whether the dynamics that govern the evolution of the series contain any predictable pattern. This constitutes a direct test of the efficient market hypothesis. An extensive dataset allows us to assess the impacts of certain periods of financial turbulence that are associated with lower asset returns and increased volatility.

First, our paper contributes to the existing literature on the predictability of option-implied measures. Specifically, we assess the predictability of the model-free implied correlation process using several specifications that capture alternative traits of the correlation distribution, namely AR(I)MA, AR(I)MA-GARCH, ARFIMA, regime-switching, heterogeneous autoregressive, and economic determinants models, as well as combination forecasts with constant and time-varying weights. Both the statistical and economic significance of the forecasting performances of constructed out-of-sample forecasts have been investigated. We find that the Schwartz combination forecasts outperform those of the competing models in terms of the accuracy of forecasts of the magnitude, whereas the equal and Schwartz combination models exhibit the strongest predictive power in terms of successful direction prediction. All of the models produce better forecasting accuracies than the random walk model that is used as the benchmark, which calls into question the informational efficiency of the S&P 100 options market.

Turning to the economic significance of the forecasts obtained, we find that the AR(I)MA-GARCH and the combination forecasts of the implied correlation are successful at generating profitable strategies. However, once transaction costs are taken into account, no economically significant profits are obtained. Our results are robust across different in-sample windows and forecast periods.

To conclude, the trading strategy implemented here does not confirm the existence of predictable patterns in the S&P 100 market, based on statistical measures, and thus, the efficiency of the S&P100 index and stock options markets cannot be rejected. Future work could investigate the profitability of correlation trading strategies further using different datasets, alternative model specifications and extended forecast horizons.

Acknowledgments

We are particularly grateful to the associate editor and two anonymous referees, whose comments have improved the paper considerably.

The research of the first author has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the operational program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF). Research funding program: Heracleitus II. Investing in knowledge society through the European Social Fund.

Appendix

Unlike the traditional, model-based Black–Scholes implied volatility, Britten-Jones and Neuberger (2000) developed a non-parametric specification of the implied variance under the assumption that the price process is continuous while the volatility is stochastic. Jiang and Tian (2005) later extended the methodology to account for the presence of jumps in the price process of the underlying asset. Specifically, these researchers suggest that under the risk-neutral measure, Q, the integrated variance between T_1 and T_2 is specified fully by a set of call options, expiring on the specified dates:

$$E^Q\left[\int_{T_1}^{T_2}\left(\frac{dS_t}{S_t}\right)^2\right] = 2\int_0^{\infty}\frac{C(T_2,K)-C(T_1,K)}{K^2}dK \approx \sum_{i=1}^{m}\frac{C^F(T,K_i)-\max(0,F_{0,T}-K_i)}{K_i^2} + \frac{C^F(T,K_{i-1})-\max(0,F_{0,T}-K_{i-1})}{K_{i-1}^2}\Delta K, \tag{25}$$

where $C^F(T,K_i)$ is the forward price of a call option, with a strike price equal to K_i, expiring at time T; S_t is the underlying asset price; $F_{0,T}$ is the forward asset price; $K_i = K_{\min} + i\Delta K$ for $0 \le i \le m$; and $\Delta K = (K_{\max} - K_{\min})/m$.

The model-free implied methodology relies on the assumption of European-style options. However, our dataset consists of the options on the S&P 100 index and on the constituent stocks, all of which are American-style, and therefore incorporates the early exercise premium (EEP). Tian (2011) proposed a methodology for extracting the risk-neutral density, and subsequently the EEP, from American options with either discrete or continuous dividends. He proposed an iterative implied binomial tree (*i*IB) as a modification of the standard implied binomial tree. The main steps of the methodology are outlined below.

First of all, we consider the observed American option price as an initial estimate of the European price. We build a standard implied binomial tree and calibrate the ending nodal probabilities to the initial option prices. Because of the limited availability of strike prices from traded options, and with the goal of maintaining the non-parametric feature of the methodology, we follow Tian (2011) and choose the ending nodal probabilities through an optimization procedure that maximizes the smoothness of the implied risk-neutral density.

Based on these nodal probabilities, we then use the implied binomial tree to price both European and American options and obtain an estimate of the early exercise premium. Having obtained the EEP, the refined estimate of a European option price is defined as the difference between the observed American price and the EEP. The refined European price is then considered as the initial input for the second iteration, in which the optimization procedure is repeated until the sum of the squared incremental changes in EEP has been minimized. We find that the average EEP for all options across the full dataset is equal to 0.51%.

Having corrected the observed American prices and obtained the corresponding European option prices, we calculate the corresponding Black–Scholes implied volatilities and proceed with the implementation of Jiang and Tian's (2005) methodology. More specifically, we collect data for each option in our dataset across moneyness levels at each point in time. In the case of multiple maturities, we select the two that are closest to the 30-day horizon. We first calculate the moneyness level for every option in our dataset.

In cases where we come across duplicate moneyness levels, i.e., call and put options with the same strike prices, we compute the average of the implied volatilities. When calculating the model-free implied volatility of an option for each day, we require at least three unique implied volatility points with a specific time-to-maturity.

In the next step, we fit a piecewise cubic Hermite interpolating polynomial to create a fine grid of implied volatilities that correspond to the range of moneyness defined earlier. For any moneyness level beyond the existing bounds, we use the last-known implied volatility value on this boundary. Thereafter, 1000 implied volatility points are translated into call prices using the Black–Scholes model, and Eq. (25) is applied.[10] As Jiang and Tian (2005) noted, the use of the Black–Scholes model does not dilute the model-free properties of their specification. The Black–Scholes model is not assumed to be the true process for the underlying price process; instead, this model is employed only as an intermediate tool for translating implied volatilities into call prices. After obtaining the model-free implied volatility for a pair of options of different maturities, we apply a cubic spline across the predetermined fixed maturity of 30 days. Finally, the model-free implied volatility is annualized on a 30/365 basis.

References

Aksu, C., & Gunter, S. I. (1992). An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, *8*, 27–43.

Becker, R., & Clements, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting*, *24*, 122–133.

Bernales, A., & Guidolin, M. (2014). Can we forecast the implied volatility surface dynamics for CBOE equity options? *Journal of Banking & Finance*, *46*, 326–342.

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

Bourgoin, F. (2001). Stress-testing correlations: An application to portfolio risk management. In C. Dunis, A. Timmermann, & J. Moody (Eds.), *Developments in forecast combination and portfolio choice*. New York: Wiley.

Britten-Jones, M., & Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, *55*, 839–866.

Brooks, C., & Oozeer, M. C. (2002). Modelling the implied volatility of options on long gilt futures. *Journal of Business, Finance and Accounting*, *29*, 111–137.

Buraschi, A., Trojani, F., & Vedolin, A. (2013). Economic uncertainty, disagreement, and credit markets. *Management Science*, *60*, 1281–1296.

Buss, A., & Vilkov, G. (2012). Measuring equity risk with option-implied correlations. *Review of Financial Studies*, *25*, 3113–3140.

Campa, J. M., & Chang, P. H. K. (1998). The forecasting ability of correlations implied in foreign exchange options. *Journal of International Money and Finance*, *17*, 855–880.

Chang, B. Y., Christoffersen, P., Jacobs, K., & Vainberg, G. (2012). Option-implied measures of equity risk. *Review of Finance*, *16*, 385–428.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174–196.

[10] Eq. (25) raises two main implementation issues. First, the trapezoid rule for numerical integration leads to discretization errors. Second, only a finite number of strike prices, $[K_{\min}, K_{\max}]$, is available for a given trading date. To overcome these limitations, we follow Jiang and Tian (2005) to create a fine grid of $m = 1000$ points with option prices within the moneyness range of [0.3, 3]. Tests using alternative numbers of integral points, m, and moneyness ranges do not produce significant changes. The results are available upon request.

Cosemans, M. (2011). *The pricing of long and short run variance and correlation risk in stock returns*. Working paper, Erasmus University–Rotterdam school of management.

David, A., & Veronesi, P. (2002). *Option prices with uncertain fundamentals*. Olin school of business working paper, 07-001.

DeMiguel, V., Plyakha, Y., Uppal, R., & Vilkov, G. (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis*, *48*, 1813–1845.

Driessen, J., Maenhout, P. J., & Vilkov, G. (2009). The price of correlation risk: evidence from equity options. *Journal of Finance*, *64*, 1377–1406.

Driessen, J., Maenhout, P.J., & Vilkov, G. (2012). *Option-implied correlations and the price of correlation risk*. Advanced risk and portfolio management paper.

Elton, E. J., & Gruber, M. J. (1973). Estimating the dependence structure of share prices–implications for portfolio selection. *The Journal of Finance*, *28*, 1203–1232.

Engle, R., & Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business and Economic Statistics*, *30*, 212–228.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1994). Forecasting international equity correlations. *Financial Analysts Journal*, *50*, 32–45.

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

Goyal, A., & Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, *94*, 310–326.

Guidolin, M., & Timmermann, A. (2003). Option prices under Bayesian learning: implied volatility dynamics and predictive densities. *Journal of Economic Dynamics and Control*, *27*, 717–769.

Guo, D. (2000). Dynamic volatility trading strategies in the currency option market. *Review of Derivatives Research*, *4*, 133–154.

Han, B. (2007). Stochastic volatilities and correlations of bond yields. *Journal of Finance*, *62*, 1491–1524.

Härdle, W.K., & Silyakova, E. (2012). *Implied basket correlation dynamics*. Working paper, Humboldt University.

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

Jiang, G. T., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *Review of Financial Studies*, *18*, 1305–1342.

Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*, *27*, 238–251.

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

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

Lopez, J. A., & Walter, C. (2000). Is implied correlation worth calculating? Evidence from foreign exchange options and historical data. *Journal of Derivatives*, *7*, 65–82.

McCracken, M. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, *140*, 719–752.

Moskowitz, T. (2003). An analysis of covariance risk and pricing anomalies. *Review of Financial Studies*, *16*, 417–457.

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

Ni, S., Pan, J., & Poteshman, A. (2008). Volatility information trading in the option market. *Journal of Finance*, *63*, 1059–1091.

Palandri, A. (2009). *The effects of interest rate movements on assets' conditional second moments*. School of economics and management Aarhus University, creates research paper, 32.

Paye, B.S. (2008). *Do macroeconomic variables predict aggregate stock market volatility?* Unpublished working paper, Rice University.

Peng, L. (2005). Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis*, *40*, 307–329.

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*, 563–602.

Peng, L., Xiong, W., & Bollerslev, T. (2007). Investor attention and time-varying comovements. *European Financial Management*, *13*, 394–422.

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, *10*, 461–465.

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

Pollet, J. M., & Wilson, M. (2010). Average correlation and stock market returns. *Journal of Financial Economics*, *96*, 364–380.

Rubinstein, M. (1976). The valuation of uncertainty income streams and the pricing of options. *Bell Journal of Economics*, *7*, 407–425.

Sheppard, K. (2008). *Economic factors and the covariance of equity returns*. Working paper, University of California, San Diego.

Skintzi, V. D., & Refenes, A. P. N. (2005). Implied correlation index: a new measure of diversification. *Journal of Futures Markets*, *25*, 171–197.
Tian, Y. S. (2011). Extracting risk-neutral density and its moments from American option prices. *Journal of Derivatives*, *18*, 17–34.

Chryssa Markopoulou is a Postdoctoral Research Fellow at the Athens University of Economics and Business (AUEB), Department of Management Science and Technology. She obtained a Ph.D. in Financial Risk Management and an M.B.A. degree with specialization in financial engineering from AUEB. For the last five years, she has been a teaching assistant of Quantitative Methods in Finance and Financial Engineering at AUEB. Her research interests lie in the area of financial econometrics and risk management, and her previous work has been presented at various international conferences.

Vasiliki Skintzi is an Assistant Professor in the Department of Economics of the University of Peloponnese, Greece. She holds a Ph.D. in Financial Engineering from the Athens University of Economics and Business, Greece, and a M.Sc. from the University of Warwick, UK. Her research interests are in financial econometrics, risk management and international financial markets. She has published in the *Journal of Financial Econometrics*, *Journal of Forecasting*, *Journal of Futures Markets*, *Journal of International Financial Markets, Institutions and Money*, and *Journal of Alternative Investments*. She has acted as a referee for academic journals and has presented her work at various international conferences.

Apostolos Refenes is Professor of Financial Engineering in the Department of Management Science and Technology and Director of the Financial Engineering Research Centre at Athens University of Business & Economics. He has previously held academic appointments at London Business School (Associate Professor) and University College London (Senior Research Fellow). He has authored over 100 papers and edited six books on the subjects of neural computing and financial engineering applications, and has been the recipient of research awards of in excess of $10m from several public and private funds. He is also associate editor of the *Intelligent Systems in Accounting and Finance Journal* and the *International Journal of Computational Intelligence and Organisations*, guest editor of the *Journal of Forecasting,* and a programme committee member for many international conferences.