Journal of Banking & Finance 37 (2013) 4943–4957

Contents lists available at ScienceDirect

Journal of Banking & Finance

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

Forecasting EUR–USD implied volatility: The case of intraday data

Christian Dunis[a], Neil M. Kellard[b,*], Stuart Snaith[b]

[a] Horus Partners Wealth Management Group SA, 1 Rue de la Rôtisserie, 1204 Genève, Switzerland
[b] Essex Business School, Essex Finance Centre, University of Essex, United Kingdom

ARTICLE INFO

Article history:
Received 6 March 2012
Accepted 31 August 2013
Available online 13 September 2013

JEL classification:
C22
C32
C53
C58
G17

Keywords:
Exchange rates
Implied volatility
Intraday data
Out-of-sample prediction

ABSTRACT

This study models and forecasts the evolution of intraday implied volatility on an underlying EUR–USD exchange rate for a number of maturities. To our knowledge we are the first to employ high frequency data in this context. This allows the construction of forecasting models that can attempt to exploit intraday seasonalities such as overnight effects. Results show that implied volatility is predictable at shorter horizons, within a given day and across the term structure. Moreover, at the conventional daily frequency, intraday seasonality effects can be used to augment the forecasting power of models. The type of inefficiency revealed suggests potentially profitable trading models.

1. Introduction

There is a large literature investigating the ability of implied volatility (hereafter IV) to predict realised volatility. Recent work includes Taylor et al. (2010) and Muzzioli (2010), and shows IV outperforms the competing model free volatility measure for US stock prices and the DAX index respectively. On the other hand and among many others, Becker et al. (2007) argue that IV has no incremental information above that offered by a combination of model based forecasts.[1] Notably, employing the same S&P 500 sample, this finding is overturned by Becker et al. (2009) after allowing for jump components in the underlying asset.

By contrast, the forecastability of IV *itself* is a relatively under-researched area. This is a surprising omission given that IV is frequently considered a proxy for market risk and subsequently, an input to asset pricing models. For example, Pojarliev and Levich (2008) examine indexed foreign exchange (FX) trader returns and the returns of individual currency managers, finding IV to be a significant explanatory factor post-2000. Hibbert et al. (2008) and Corrado and Miller (2006) also assess the IV-return relationship but for the S&P 500; the former arguing that a behavioural framework provides a rationale for the empirical results, and the latter that IV yields a superior predictor of realised excess returns. As a corollary, predictability in IV should allow a richer understanding of the dynamics of expected returns. Moreover, given IV is defined in relation to a relevant option price, some degree of IV predictability may allow the construction of profitable option trading strategies (see Konstantinidi et al., 2008) and prompt questions regarding option market efficiency.

The relatively small number of extant empirical studies examining IV predictability typically focus on equity or equity indices. Adopting a number of forecasting model specifications (e.g., relevant economic variables, time series models and principal components) Konstantinidi et al. (2008) examine several US and European stock IV indices.[2] Dumas et al. (1998), Gonçalves and Guidolin (2006) and Andreou et al. (2010) assess predictability of the IV surface on the underlying S&P 500, whilst analogously, Chalamandaris and Tsekrekos (2010) consider options from several foreign exchange rates. Earlier work includes Harvey and Whaley

* Corresponding author. Address: Essex Business School, Essex Finance Centre, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom. Tel.: +44 1206 874153; fax: +44 1206 873429.
E-mail address: nkellard@essex.ac.uk (N.M. Kellard).

[1] For a very useful survey article see Poon and Granger (2003). Jumps are also addressed by Busch et al. (2011) for both foreign exchange and the S&P 500.

[2] For other research, see Konstantinidi and Skiadopoulos (2011) for an examination of the forecastability of VIX futures prices and Kim and Kim (2003) for a study investigating the dynamics of IV from currency options on futures.

0378-4266/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jbankfin.2013.08.028

(1992) who employ at-the-money (ATM) IV on S&P 100. Overall, the studies cited above tend to uncover evidence of predictability in IV that is not subsequently translated into economic significance.

This paper adds to the existing literature by examining the forecastability of high frequency IV. Our work is closest in spirit to Konstantinidi et al. (2008), insofar as whilst the markets examined are different, both apply a variety of forecasting models to try and predict the evolution of IV. However, the paper is distinct from Konstantinidi et al. (2008) and the extant literature in a number of ways, making several contributions to the literature. Firstly, to our knowledge, this is the first intraday study of IV predictability – using a novel dataset sampled at a 30-min frequency for the EUR–USD exchange rate, with data available at a number of maturities across the term structure. Secondly, whether IV is predictable at the intraday frequency (i.e., at short horizons, within a single day) is evaluated by comparing a variety of forecasting models against a random walk benchmark. Thirdly, in a new test of market efficiency, we assess whether intraday data and related seasonalities, such as overnight (ON) effects, are useful in outperforming or augmenting forecasting models at the conventional daily frequency. For this two other comparator datasets are derived from the high frequency data: a half-daily sampled dataset allowing the assessment of intraday seasonalities and a daily sampled dataset to represent the conventional frequency. Fourthly, models are compared out-of-sample by applying (i) Hansen's (2005) test for superior predictive ability (SPA) to test whether models are outperformed by alternative forecasts (ii) Hansen et al.'s (2011) model confidence set (MCS) to return a subset of 'best' models with a given level of confidence (iii) Pesaran and Timmermann's (1992) predictive failure test, which examines the ability of models to predict the direction of IV, and a pseudo trading strategy, both of which are reflective of likely trading success.[3]

The preliminary in-sample results are interesting. Among other things we find that mean reversion dynamics appear stronger in the 30-min dataset than a daily comparator, that GARCH models can be used to model the 'volatility' of volatility and that diurnal seasonality patterns such as ON effects appear significant in our regressions. Can these revealed effects contribute to improved out-of-sample IV forecasting performance? Strikingly, we find strong evidence that time series models using the 30-min data outperform the random walk benchmark up to 5 hours ahead. Clearly useful information for market participants trading intraday, but interestingly this performance disappears when considering one-day ahead forecasts. However, it is shown that incorporating ON effects yields models that can outperform both in terms of forecasting the directional change in IV and in our pseudo trading exercise. In particular, ON models capture a 'weekend effect' (i.e., a typically low late Friday value of IV, compared with a high Monday morning value) which allows particularly improved performance when forecasting Friday and Monday IV and signals the existence of a type of market inefficiency.

The remainder of this paper is structured as follows. Section 2 presents the dataset and Section 3 details the time series models to be employed. Sections 4 and 5 present the in-sample and the out-of-sample results respectively. A final section concludes.

2. Data

Throughout the paper, intraday IV data on the EUR–USD exchange rate are employed for the period November 2004 to September 2008.[4] As in Kellard et al. (2010) and Covrig and Low (2003), IV is measured by ATM, over-the-counter (OTC) market quoted volatilities for European options. The use of market quoted volatilities is different from the majority of studies in the literature whom employ implied volatility backed out from exchange-traded option prices. As Covrig and Low (2003) describe, although participants in exchange-traded markets quote prices in terms of the familiar option premium, OTC prices are given in terms of volatility. In other words, an option could be quoted at 12% p.a. and subsequently converted into the appropriate option premium by using the Garman–Kohlhagen model.

Table 1
Summary statistics on 30-min $\Delta IV_{m,t}$.

	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Mean	0.0000	0.0000	0.0000	0.0000
Median	0.0000	0.0000	0.0000	0.0000
Maximum	0.7175	0.2048	0.4055	0.4183
Minimum	−0.6931	−0.2048	−0.4079	−0.4055
Std. Dev.	0.0105	0.0072	0.0074	0.0065
Skewness	1.0383	0.0553	−0.0539	0.5349
Kurtosis	890.1177	55.7344	423.5964	886.1469
Jarque–Bera (JB)	1.58E + 09	5.57E + 06	3.55E + 08	1.56E + 09
JB (*p*-value)	0.0000	0.0000	0.0000	0.0000
ADF (*p*-value)	0.0001	0.0001	0.0001	0.0001
MZ_t	−1.84*	−100.00***	−54.08***	−41.047***
Sum	0.3243	0.2377	0.1548	0.1136
Sum Sq. Dev.	5.2754	2.4966	2.6142	2.0158
Observations	48,112	48,112	48,112	48,112

Notes: IV data are logged and first differenced. The ADF test employed an intercept (and no trend) and used SIC to select the appropriate lag length (from a maximum length of 48). The MZ_t test of Ng and Perron (2001) uses GLS detrending and also selects lag length in the former manner.
* Significance at 10% level.
** Significance at 5% level.
*** Significance at 1% level.

Given currency volatility has become a traded quantity in financial markets, it is therefore directly observable on the marketplace and the use of these volatilities avoids the potential biases (i.e., errors in the choice of option pricing model and the measurement of model inputs) associated with backing out data from an option pricing model.[5] In any case, the OTC FX market is vastly more liquid than its exchange-traded counterpart. For example, at the end of June 2012, the Bank of International Settlement (2012) reported the notional amount outstanding in the OTC currency option market stood at $11.1 trillion, compared with $111 billion for the exchange-traded market. Moreover, the US Dollar (i.e., $8.7 trillion outstanding) and the Euro (i.e., $4.1 trillion outstanding) are the two most heavily traded currencies within the option OTC market.

The data employed in this study are at a 30 min frequency from 12 am Monday to 11.30 pm Friday (London time) for 1, 3, 6 and 12-month maturities yielding a raw dataset of approximately 48,000 time series observations.[6] The natural logarithms of all volatility series were taken to minimise the possibility of non-normal variables as shown by, *inter alia*, Christensen and Hansen (2002). Analogously to Konstantinidi et al. (2008), unit root tests on the level of logged IV cannot reject the null of non-stationarity.[7] Therefore, to deal with non-stationary in the IV level we forecast the change in IV, $\Delta IV_{m,t}$ with maturity m at time t. Summary statistics and

[3] We are grateful to Peter Hansen and Asger Lunde for making the code for the SPA and MCS tests freely available to run in Ox.

[4] The start and end dates were governed by the availability of the intraday IV data from Reuters feeds.

[5] Implied volatilities are also annualised rates so that a quoted volatility of 5 per cent typically translates to a monthly variance rate of $(0.05^2)(21/252)$. The calculations assume that annualised rates refer to a 252 trading day year.

[6] The global foreign exchange currency spot market typically trades continuously from 10 pm Sunday (Sydney open) to 10 pm Friday London time (New York close). The 11.30 pm quotes on Friday occasionally differed from the 10 pm value, but yielded no significant differences in model selection or parameter estimation. Unfortunately, assuming they exist, the dataset did not give Sunday IV values. All presented results in this paper are based on the full available dataset.

[7] Results are not tabulated here to save space but can be provided on request.

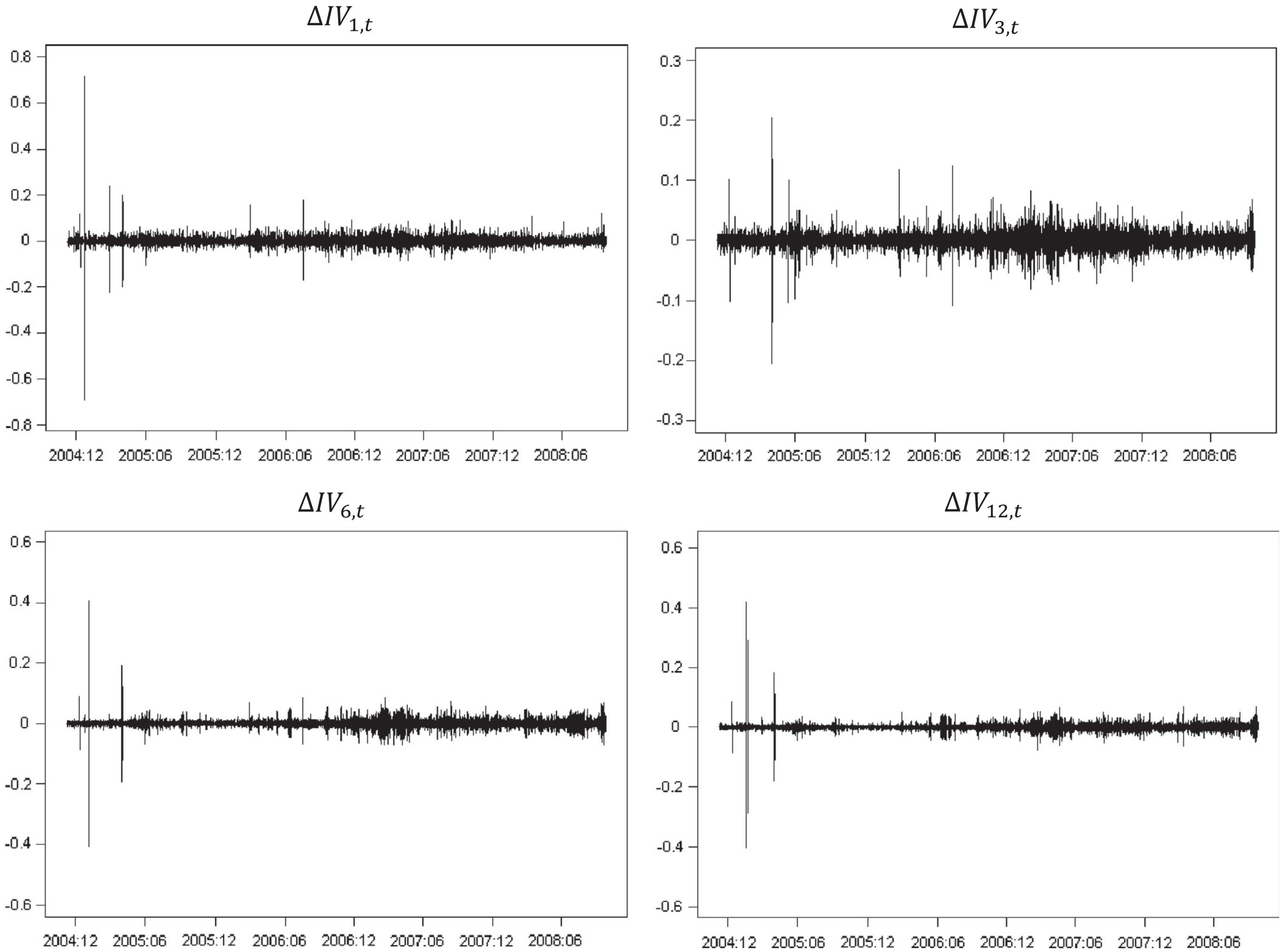

Fig. 1. Intraday $\Delta IV_{m,t}$ (10/11/2004 to 25/09/2008). *Notes*: $\Delta IV_{1,t}$ corresponds to the 30-min first difference of the logged 1-month IV series. Analogous representation is given for the 3, 6 and 12-month series.

time-series plots for the data can be found in Table 1 and Fig. 1 respectively.

Notably, the unit root tests in Table 1 suggest that for each maturity, $\Delta IV_{m,t}$ is a stationary variable. To provide a preliminary assessment of any intraday seasonality and to observe the trajectory of IV during the trading day we examined the average raw IV value at each 30 min observation for each day of the week. Fig. 2 summarises.

Most obviously there is a possible *weekend effect* across the maturities, with a low average late Friday value being contrasted with a high average early Monday morning value. To assess this further we construct the average half-daily logged returns for implied volatility where DAY represents the average 4 am to 4 pm return and ON shows the average 4 pm to 4 am return.

Supporting the volatility weekend effect findings of Fig. 2, the stand out value in Fig. 3 is Friday's ON at the 1-month maturity. Additionally, it is noticeable that the 1-month maturity produces larger returns than those of a longer maturity and that ON values are typically larger than the comparable DAY values. We shall return to these observations later in the paper.

To facilitate a comparison with the extant literature we also employ a conventional daily dataset. Specifically, the relevant time series is constructed by sampling the observed IV at 4 pm each day. Additionally, we also generate a dataset of half daily frequency, sampled at 4 am and 4 pm each day. The intraday seasonality observed in Figs. 2 and 3 motivate our data frequency; volatility returns often appearing highest overnight, with the weekend effect described above the most conspicuous example.

Finally, for the economic forecasting model we use the following additional data: daily closing EUR–USD spot prices, the price of Brent Crude Oil, and the Europe and USD interbank offered rate for 1, 3, 6, and 12 month maturities. All these data are downloaded from DataStream.

3. Empirical methodology

3.1. AR, VAR, and VECM models

Given the use of high frequency data, time series models are an appropriate choice to assess forecastability. Following Konstantinidi et al. (2008) we initially employ univariate autoregressive (AR) and vector AR (VAR) models. In the case of the former this provides a simple predictive model for the evolution of IV. In the case of the latter the multivariate structure permits comment on whether $\Delta IV_{m,t}$ can be forecasted using implied volatilities from different maturities. In both cases the maximum number of lags considered is 48 for the 30-min dataset, corresponding to a 24 h window, where the Schwarz information criterion (SIC) is used to determine the optimal lag structure Q. For the daily dataset Q is set equal to 10 and the AR specification is given by:

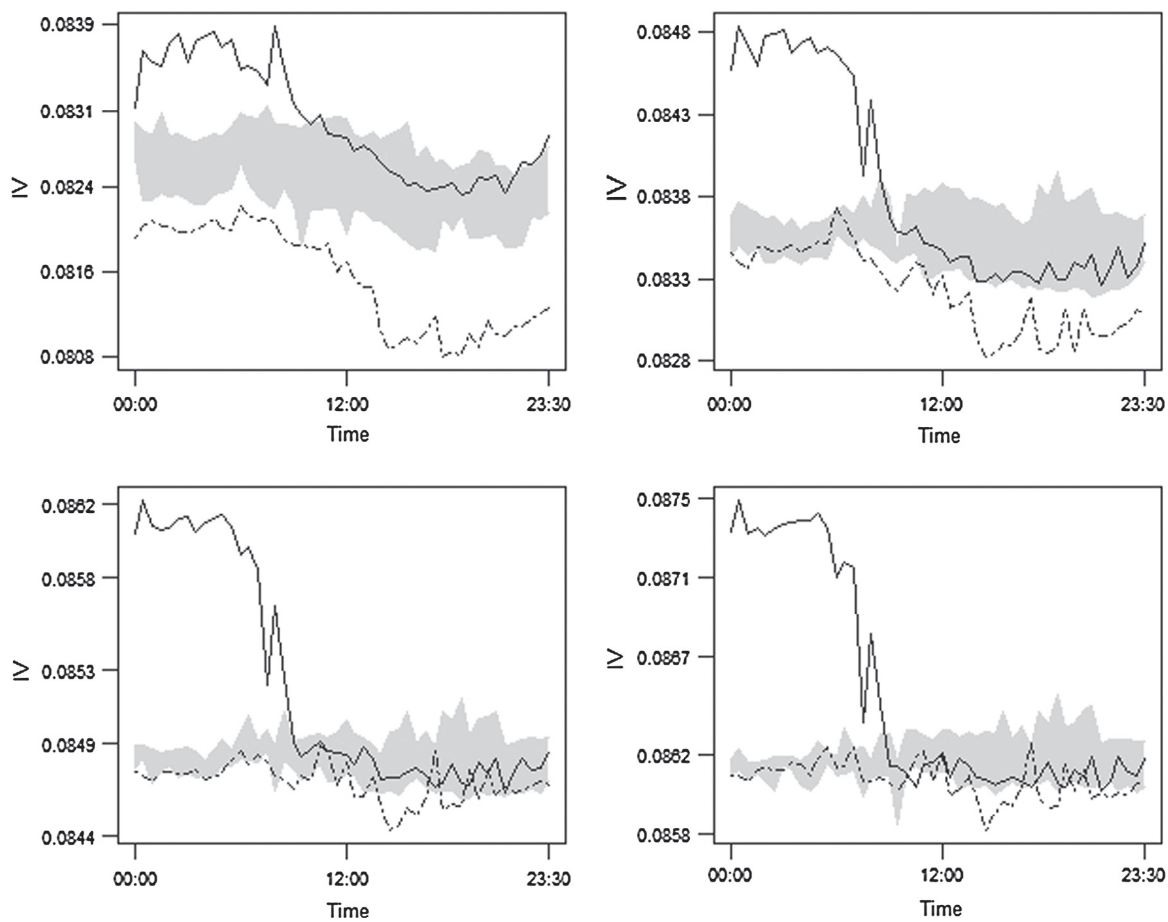

Fig. 2. Intraday raw IV dynamics. *Notes*: This figure plots the average trajectory of raw IV from 00:00 to 23:30 for Monday (black line) and Friday (dashed line) across the sample. For ease of interpretation the range at each point in the day is represented by the shaded area for the remaining days of the week. At any fixed point in time in the trading day the wider the shaded area (in the *y*-plane) the more dispersed the plots for those days of the week.

$$\Delta IV_{m,t} = \alpha + \sum_{i=1}^{Q} \beta_i \Delta IV_{m,t-i} + \varepsilon_t \tag{1}$$

where ε_t is an error term. The VAR specification is given by:

$$Y_t = A + \sum_{i=1}^{Q} B_i Y_{t-i} + \varepsilon_t \tag{2}$$

where Y_t is the vector of $\Delta IV_{m,t}$ for m = 1, 3, 6, 12 and A and B_i represent a vector of constants and a matrix of coefficients respectively. We also tested whether there is long-run relationship between IV in levels. Johansen test results (not reported but available on request) yield evidence of cointegration amongst the maturities and thus we also model a VECM model:

$$\Delta V_t = C + \Pi V_{t-1} + \sum_{i=1}^{Q} \Gamma \Delta V_{t-i} + \varepsilon_t \tag{3}$$

where V_t is the vector of $IV_{m,t}$ for m = 1, 3, 6, 12, C represents a vector of constants and Π represents the long-run coefficient matrix. Again Q is determined by SIC.

3.2. ARFIMA model

Several papers in the literature have proposed that asset price volatility is neither an I(1) nor an I(0) process but rather a fractionally integrated or I(d) process (for a discussion see Kellard et al., 2010). The introduction of the autoregressive fractionally integrated moving average (ARFIMA) model by Granger and Joyeux (1980) and Hosking (1981) allows the modelling of persistence or long memory where $0 < d < 1$. Therefore, as with Konstantinidi et al. (2008), we apply the following ARFIMA (p,d,q) model:

$$\Phi(L)(1-L)^d \Delta IV_{m,t} = c + \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim idd(0, \sigma^2) \tag{4}$$

where $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_P L^p$ and $\Theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$. For $0 < d < 0.5$, the process exhibits stationarity with long memory meaning shocks disappear hyperbolically rather than geometrically. On the other hand, when $0.5 < d < 1$, the process is mean reverting but non-stationary, presenting an unconditional variance that grows at a more gradual rate than when $d = 1$. To avoid over-fitting the data, for each maturity an ARFIMA (1, d, 1) is selected and estimated by exact maximum likelihood (EML).

3.3. AR-GARCH, AR-TGARCH and VAR-BEKK models

In an extension to the models applied thus far in the literature, we augment the above AR and VAR models by determining whether IV itself is conditionally heteroskedastic. Given the volatility trading strategies where the intertemporal difference in volatility translates to an investor's return, there is an *a priori* logic in assessing the 'volatility' of volatility. Specifically the AR(Q) model (1) above is augmented to include GARCH(1,1) yielding the following conditional variance equation:

$$\sigma_t^2 = c + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2 \tag{5}$$

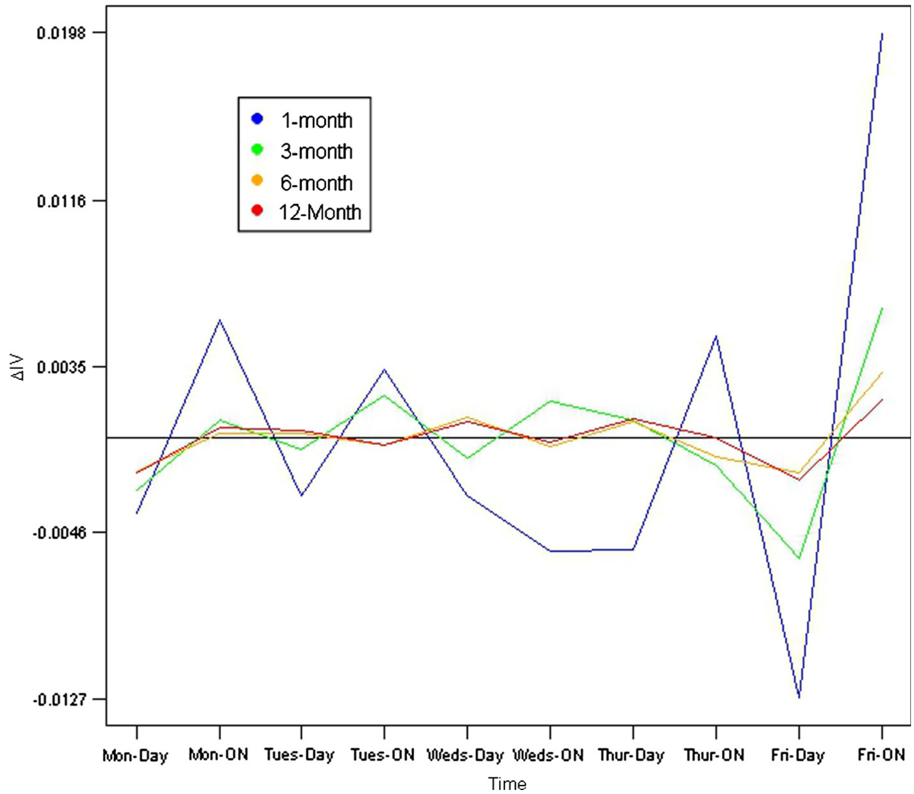

Fig. 3. 12 hour Changes in the logarithm of IV. *Notes*: This figure plots the average half-daily (i.e., 12 h) logged returns for implied volatility. DAY represents the average 4 am to 4 pm return, whilst ON shows the average 4 pm to 4 am return for each day of the week.

where γ_1 and γ_2 are the coefficients on the ARCH and GARCH terms respectively. We also implement an AR-TGARCH model which does not imposes a symmetric response of volatility to positive and negative shocks. The conditional variance of the TGARCH model is given by:

$$\sigma_t^2 = c + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2 + \gamma_3 \varepsilon_{t-1}^2 I_{t-1} \tag{6}$$

where $I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$ and zero otherwise. Finally we estimate a multivariate GARCH model where the conditional mean equation is a VAR(1) and the conditional variance–covariance matrix follows the parameterisation of Baba–Engle–Kraft–Kroner (BEKK) defined in Engle and Kroner (1995). Specifically we implement the diagonal BEKK model with one ARCH and one GARCH parameter and thus write the conditional covariance matrix H_t as:

$$H_t = C'C + A_1' \varepsilon_{t-1} \varepsilon_{t-1}' A_1 + G_1' H_{t-1} G_1 \tag{7}$$

where the A and G are diagonal matrices and C is upper triangular and all of dimension 4×4.[8]

3.4. Economic variables model

Using the daily frequency dataset we implement an economic model in the spirit of Konstantinidi et al. (2008). Of course, at this relatively high frequency, certain economic variables that might be theoretically appropriate (e.g., GDP growth) are unavailable and thus omitted. We therefore consider the following model:

$$\Delta IV_{m,t} = \alpha + \beta_1 \Delta oil_{t-1} + \beta_2 (i_{m,t-1} - i^*_{m,t-1}) + \beta_3 \Delta HV_{t-1} + \beta_4 \Delta IV^+_{m,t-1} + \beta_5 \Delta IV^-_{m,t-1} + \beta_6 \Delta s^+_{t-1} + \beta_7 \Delta s^-_{t-1} + u_t \tag{8}$$

where Δoil_{t-1} denotes the change in the Brent Crude oil price and $i_{m,t-1} - i^*_{m,t-1}$ is the *m*-period interest rate differential between the Eurozone and the United States. $\Delta IV^+_{m,t-1}$ and Δs^+_{t-1} denote positive changes in lagged IV and the underlying spot price respectively with analogous variables for negative changes. Finally ΔHV_{t-1} denotes the change of 30-day historical volatility. As with Konstantinidi et al., the choice of variables is motivated by the literature on the predictability of asset returns, to which the evolution of IV is inextricably linked. For example, we justify the inclusion of the change in oil price and the interest rate differential given their links to the underlying asset, the spot exchange rate. In the case of the former, the oil price represents probably the most important global commodity; clearly this might conceivably impact the spot exchange rate and subsequently influence IV.[9] In the case of the latter, the interest rate differential is a proxy for the market expectation of the change in the underlying spot rate under uncovered interest rate parity.

3.5. Principal components model

Principal components (PC) analysis models the variance structure of a set of variables using linear combinations of those variables. Rather than trying to model this structure completely, using as many factors as variables, it is common to choose fewer components than variables. As Konstantinidi et al. (2008) comment, PC is appealing, being shown to consistently estimate the true latent factors under quite general conditions. Given we have fewer variables than in many applications of PC, we adopt two PC out of a possible maximum of four:

[8] Estimation is via quasi-maximum likelihood and is implemented using G@RCH 6.1 on the Oxmetrics platform.

[9] Indeed recent work by Ferraro et al. (2012) has employed oil prices to successfully forecast exchange rates for a small selection of developed economies.

Table 2
Overnight day-of-the-week regression – AR-GARCH.

	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
α	−0.0055	−0.0012	−0.0002	0.0000
	(0.0007)***	(0.0004)***	(0.0003)	(0.0002)
β_1	0.0103	0.0004	0.0000	−0.0008
	(0.0019)***	(0.0013)	(0.0009)	(0.0006)
β_2	0.0080	0.0028	−0.0004	−0.0003
	(0.0024)***	(0.0014)**	(0.0011)	(0.0006)
β_3	−0.0021	0.0029	−0.0004	−0.0002
	(0.002)	(0.0013)**	(0.0011)	(0.0007)
β_4	0.0105	−0.0014	−0.0008	−0.0010
	(0.0023)***	(0.0017)	(0.0010)	(0.0007)
β_5	0.0277	0.0076	0.0029	0.0017
	(0.0015)***	(0.0011)***	(0.0009)***	(0.0006)***
δ_1	−0.1688	−0.2089	−0.2080	−0.2679
	(0.0271)***	(0.0308)***	(0.0339)***	(0.0342)***
c	0.0001	0.0000	0.0000	0.0000
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
γ_1	0.1434	0.1850	0.1435	0.1738
	(0.01500)***	(0.017)***	(0.0141)***	(0.0135)***
γ_2	0.7504	0.7545	0.8245	0.8232
	(0.0265)***	(0.0172)***	(0.0133)***	(0.0103)***
Adj. R^2	0.1681	0.0465	0.0085	−0.0067

Notes: Figures in parentheses are standard errors. β_1 to β_5 are the Monday to Friday coefficients on the ON day-of-the-week dummy variables, e.g. β_1 is the coefficient on the 4 pm Monday to 4 am Tuesday dummy variable. γ_1 and γ_2 are the coefficients on the ARCH and GARCH terms respectively.
* Rejection of zero null at 10%.
** Rejection of zero null at 5%.
*** Rejection of zero null at 1%.

$$\Delta IV_{m,t} = \alpha + \beta_1 PC1_{t-1} + \beta_2 PC2_{t-1} + u_t \tag{9}$$

This model has to be further adapted in the intraday case to implement multi-step ahead forecasting. In this case each principal component is forecast by an AR-process, and those forecasts in turn used to generate conditional multi-step forecasts for $\Delta IV_{m,t}$.

The PC model is able to describe 71% (93%) of the total in-sample variance of the intraday (daily) changes in implied volatility across maturities. For both frequencies of data the PC1's loadings are positive and evenly distributed across maturity. However, PC2 has a large positive loading for the 1-month maturity series relative to the smaller and often negative loadings for the remaining maturities. Two PC's were chosen to introduce a parsimonious alterative to some of the more complex models we apply whilst still describing a respectable proportion of the total variance across maturities.

3.6. Day-of-the-week and overnight models

Using the half-daily frequency we can examine seasonality in the data. More specifically, volatility returns from 4 am till 4 pm (daytime) and from 4 pm till 4 am (overnight) can be dichotomised by the inclusion of overnight (ON) dummy variables for each day of the week. The simple ON model is given by:

$$\Delta IV_{m,t} = \alpha + \beta_1 D_{ON,Mon} + \beta_2 D_{ON,Tue} + \beta_3 D_{ON,Wed} + \beta_4 D_{ON,Thu} + \beta_5 D_{ON,Fri} + u_t \tag{10}$$

where α represents the mean daytime return, $D_{ON,Mon}, \ldots, D_{ON,Fri}$ are the overnight dummy variables for each day of the week, where $D_{ON,J} = 1$ overnight on day J and 0 elsewhere, and therefore the corresponding average overnight return is given by $\beta_1, \ldots, \beta_5$. In addition to this simple dummy variable model, we can account for serial correlation and a potentially non-constant variance. In the case of the former this is achieved by the inclusion AR terms in the mean equation, and in the latter by modelling both GARCH and TGARCH errors:

$$\Delta IV_{m,t} = \alpha + \beta_1 D_{ON,Mon} + \beta_2 D_{ON,Tue} + \beta_3 D_{ON,Wed} + \beta_4 D_{ON,Thu} + \beta_5 D_{ON,Fri} + \sum_{i=1}^{Q} \delta_i \Delta IV_{m,t-i} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \tag{11}$$

where a maximum lag of 20 is chosen for Q to match the window used for daily data and σ_t^2 is defined as per Eqs. (5) and (6) for GARCH and TGARCH models respectively.[10]

4. In-sample results

4.1. Daily in-sample results

To make an initial comparison with the extant literature, we estimate the applicable models from the previous section using daily data and an estimation window from 10/11/2004–25/09/2007. To conserve space these in-sample results are available in Panel A1 of Appendix A.

In summary, we find that models containing autoregressive components show some evidence of mean reversion, with the ARFIMA model suggesting anti-persistence. Comparing the in-sample fit of the VAR and VECM models, it appears that modelling the long-run relationship yields a better in-sample fit as indicated by the higher adjusted R^2 values, whereas by contrast the PC model does no tend to yield significant coefficients. Interestingly, the AR-GARCH model shows a persistent 'volatility' of volatility effect in the daily data, evidenced by volatility clustering and forecastability, with the AR-TGARCH providing little support for asymmetric effects.

Finally, the economic model yields some interesting results. This model shows that for shorter maturities, changes in the price of oil and negative changes in the underlying spot rate may be useful in forecasting changes in EUR–USD IV.

4.2. Half-daily in-sample results

Can any intraday seasonality usefully augment the information captured by the daily models above? To begin answering that question Table 2 shows the in-sample results of the ON-GARCH model, employing the same in-sample estimation window as in the daily case (i.e., 10/11/2004–25/09/2007).

The Friday ON dummy variable is positive and significant across all maturities. This clearly captures the observed behaviour of IV over the weekend, being indicative of the return from holding IV from Friday night to Monday morning being in excess of that from holding over the weekly 'daytime' (4 am–4 pm) average. It is also interesting to note that a far higher adjusted R^2 is presented at the 1-month horizon (16.81%), versus the next highest of 4.65%. In terms of the variance equation the ON-GARCH model predictably (given the daily results presented previously) yields evidence of 'volatility of volatility' clustering.[11]

4.3. 30-min in-sample results

We now turn to the highest frequency data available to us which again, to conserve space, are summarised in Panel A2 of Appendix A. Note that we use a truncated in-sample estimation window (i.e., 10/11/2004–10/1/2005 which covers the initial 2112 observations) as compared to the daily or half-daily regres-

[10] As Doyle and Chen (2009) comment, until fairly recently a simple dummy variable regression was the standard way to analyse weekday effects. More recently more complex models are used, including GARCH errors and augmentations to the mean equation.

[11] To conserve space the simple ON model and ON-TGARCH in-sample results are not presented. They likewise exhibit significant Friday dummy variable coefficients across all maturities. Full results are available from the Authors on request.

Table 3
30-min Out-of-sample forecasts versus the random walk model – 1-month maturity.

Model	h	1	2	3	4	5	6	7	8	9	10
Random walk	MSE	2.3070	3.3990	4.0387	4.5878	5.3308	5.7752	6.2772	6.8493	7.4966	8.0308
	MAE	79.6848	105.8153	122.5603	133.8050	149.1535	157.6218	166.2836	176.0584	186.4145	192.9616
AR	MSE	0.8215	1.1891	1.4745	1.7457	2.0302	2.2743	2.5245	2.7870	3.0474	3.2941
	MAE	49.2916	63.6201	73.1063	80.9433	89.3469	95.3884	101.2751	107.6100	113.5757	118.6410
ARFIMA	MSE	0.8184	1.1841	1.4707	1.7433	2.0279	2.2706	2.5226	2.7840	3.0420	3.2858
	MAE	49.4750	64.1784	74.0923	82.1530	90.5869	96.7113	102.7283	109.0093	114.9256	119.8670
VAR	MSE	0.8401	1.1885	1.4760	1.7473	2.0325	2.2788	2.5327	2.7975	3.0625	3.3089
	MAE	49.7462	63.5287	72.6566	80.2271	88.5251	94.8732	101.0093	107.3653	113.4989	118.5727
VECM	MSE	0.8380	1.1894	1.4821	1.7588	2.0527	2.3035	2.5636	2.8344	3.1034	3.3521
	MAE	50.4500	64.6800	74.2900	82.3100	91.1400	97.6600	103.9200	110.3900	116.5400	121.6300
GARCH	MSE	0.8110	1.1819	1.4732	1.7497	2.0415	2.2977	2.5640	2.8483	3.1394	3.4253
	MAE	48.9854	63.4819	73.2619	81.3863	89.9537	96.3689	102.5595	109.2740	115.5381	121.1170
TGARCH	MSE	0.8106	1.1812	1.4719	1.7480	2.0373	2.2885	2.5488	2.8272	3.1062	3.3771
	MAE	48.9217	63.4145	73.2249	81.2996	89.8177	96.1447	102.2544	108.9068	115.0314	120.4886
VAR-BEKK	MSE	1.0285	1.3448	1.6740	1.7955	2.3965	2.6080	2.8797	3.2870	3.1597	3.5363
	MAE	52.7364	64.5757	75.9148	80.4528	94.8506	100.6538	106.9770	115.8608	114.9561	122.6709
PC	MSE	0.8769	1.2944	1.6004	1.8770	2.1807	2.4155	2.6583	2.9369	3.2083	3.4549
	MAE	47.3762	62.5442	73.1396	81.0780	90.3524	96.4036	102.3207	109.1884	115.6442	120.6037

Notes: To aid interpretation all values have been multiplied by 10^4. h represents the size of the step ahead returns forecast. Mean squared error (MSE) and mean absolute error (MAE) shown for 1–10 step ahead forecasts. For h = 1–10, each model is tested for equal forecast predictability against the random walk using the (i) Diebold–Mariano (1995) test (MSE and MAE) (ii) the Clark and West (2007) mean squared prediction error test. In all cases the null hypothesis of equal forecast predictability is rejected in favour of the model at the 1% significance level.

Table 4
30-min Out-of-sample forecasts versus the random walk model – 3-month maturity.

Model	h	1	2	3	4	5	6	7	8	9	10
Random walk	MSE	1.3749	1.9836	2.2989	2.5204	2.9092	3.1015	3.3605	3.6264	3.9387	4.1274
	MAE	60.2155	79.1033	89.8779	96.3567	106.8400	111.8248	117.2608	123.4717	129.8205	132.7132
AR	MSE	0.4689	0.6535	0.7915	0.9162	1.0530	1.1702	1.2935	1.4190	1.5431	1.6550
	MAE	36.5653	46.1871	52.2138	56.9502	62.1705	65.8663	69.7249	73.7488	77.4258	80.3653
ARFIMA	MSE	0.4683	0.6515	0.7902	0.9143	1.0512	1.1683	1.2910	1.4130	1.5352	1.6448
	MAE	37.1537	47.1107	53.4248	58.2821	63.5290	67.2209	71.0524	74.8403	78.4435	81.2840
VAR	MSE	0.4730	0.6489	0.7880	0.9090	1.0476	1.1638	1.2877	1.4118	1.5359	1.6471
	MAE	37.3485	46.6421	52.4569	56.9917	62.0598	65.8280	69.7840	73.7235	77.4190	80.3592
VECM	MSE	0.4721	0.6489	0.7898	0.9141	1.0610	1.1852	1.3183	1.4482	1.5792	1.6976
	MAE	38.1402	47.8169	54.0583	59.0468	64.6625	68.7493	72.9196	76.9842	80.7819	83.9447
GARCH	MSE	0.4664	0.6532	0.7883	0.9141	1.0508	1.1688	1.2918	1.4179	1.5417	1.6554
	MAE	36.5762	46.2736	52.3248	57.1855	62.3458	66.0887	69.9380	73.9897	77.6607	80.6850
TGARCH	MSE	0.4663	0.6529	0.7885	0.9145	1.0520	1.1712	1.2947	1.4216	1.5460	1.6605
	MAE	36.5139	46.2987	52.4152	57.3325	62.5509	66.3377	70.2165	74.2912	78.0055	81.0704
VAR-BEKK	MSE	0.6162	0.7684	0.9365	0.9546	1.2822	1.3747	1.5208	1.7345	1.6104	1.7871
	MAE	40.4641	48.2272	55.7015	57.3893	67.8524	71.0062	75.0894	81.2141	79.2207	84.0554
PC	MSE	0.5159	0.7342	0.8796	1.0033	1.1527	1.2663	1.3876	1.5164	1.6456	1.7483
	MAE	35.4212	46.0818	52.8386	57.6253	63.7739	67.3741	70.9830	75.3570	79.3982	81.9999

Notes: To aid interpretation all values have been multiplied by 10^4. h represents the size of the step ahead returns forecast. Mean squared error (MSE) and mean absolute error (MAE) shown for 1–10 step ahead forecasts. For h = 1–10, each model is tested for equal forecast predictability against the random walk using the (i) Diebold–Mariano (1995) test (MSE and MAE) (ii) the Clark and West (2007) mean squared prediction error test. In all cases the null hypothesis of equal forecast predictability is rejected in favour of the model at the 1% significance level.

sions. This is to keep the number of observations for the 30-min frequency dataset computationally manageable during the later forecasting exercise.

The results from the AR, VAR and VAR-BEKK models present stronger evidence, relative to daily data, of mean-reversion and predictability in the dynamics of $\Delta IV_{m,t}$ across all maturities. To reinforce this notion note the R^2 values for the intraday AR and VAR models are noticeably higher than their daily counterparts. Again, there is some evidence that term structure might help describe IV, as evidenced by the significant non-diagonal lags in the VAR model. Unlike in the daily case, the VECM model does not yield much improvement in the adjusted R^2. For the ARFIMA model the long-memory parameter is negative and significant for the 3 and 12-month maturities, whilst for the 1 and 6-month horizons it is insignificant. Thus, unlike the daily case, we are unable to find evidence of anti-persistence in $\Delta IV_{m,t}$ across all maturities. It is worth noting that Caporale and Gil-Alana (2010a,b) demonstrate that the fractional order of integration can be affected by data frequency; specifically that a lower order of frequency can be associated with a lower order of integration.

Turning to the AR-GARCH and AR-TGARCH models it is clear at the intraday level evidence for ARCH-type effects appears ambiguous.[12] There are a number of cases where estimates for the conditional variance equation violate the necessary non-negativity constraints, although this is by no means uniform. Overall, although the ARCH-type models are possibly a misspecification at our intraday frequency we include them in the following analysis for comparative purposes (see Andersen and Bollerslev, 1997). Finally, the PC model yields more significant coefficients than its daily

[12] Using conventional price returns, instead of the returns to volatility in this current paper, Locke and Sayers (1993) also show that at the intraday frequency, evidence for ARCH effects is mixed. Where ARCH effects are present Engle and Sokalska (2012) and others note that diurnal patterns may affect coefficient estimates and measures of persistence.

Table 5
30-min Out-of-sample forecasts versus the random walk model – 6-month maturity.

Model	h	1	2	3	4	5	6	7	8	9	10
Random walk	MSE	1.3194	1.7575	2.0326	2.1073	2.4045	2.4722	2.6740	2.7845	3.0325	3.0901
	MAE	54.1950	68.5984	77.7703	81.2081	90.2210	92.6016	97.0575	100.6758	106.0634	107.2241
AR	MSE	0.4106	0.5295	0.6194	0.6923	0.7823	0.8530	0.9321	1.0080	1.0903	1.1592
	MAE	33.0048	39.9508	44.3360	47.3417	51.1010	53.4076	56.0525	58.6116	61.1834	63.1477
ARFIMA	MSE	0.4087	0.5272	0.6170	0.6905	0.7797	0.8498	0.9294	1.0067	1.0879	1.1557
	MAE	33.3865	40.5333	45.1621	48.3415	52.0306	54.3834	57.0733	59.6313	62.2205	64.1252
VAR	MSE	0.4072	0.5247	0.6174	0.6914	0.7826	0.8508	0.9292	1.0051	1.0870	1.1552
	MAE	33.2460	39.7701	44.0030	46.9376	50.6749	53.2545	56.0374	58.5877	61.2683	63.1487
VECM	MSE	0.4168	0.5267	0.6230	0.7016	0.8035	0.8734	0.9605	1.0393	1.1261	1.1942
	MAE	34.6549	41.6803	46.3557	49.7084	54.0439	56.8088	59.7664	62.2493	64.8966	66.8533
GARCH	MSE	0.4120	0.5370	0.6302	0.7015	0.7912	0.8626	0.9457	1.0187	1.1016	1.1720
	MAE	33.0636	40.0637	44.7730	47.8094	51.6557	54.0416	56.8658	59.4413	62.1069	64.1046
TGARCH	MSE	0.4068	0.5290	0.6214	0.6949	0.7813	0.8525	0.9319	1.0063	1.0890	1.1599
	MAE	32.6247	39.7272	44.2614	47.4513	51.1758	53.6390	56.3362	59.0140	61.6325	63.7476
VAR-BEKK	MSE	0.5799	0.6629	0.8167	0.7641	1.0975	1.1315	1.4695	1.8518	2.7845†	5.3656†
	MAE	37.2696	41.7647	48.9292	48.3048	58.7504	59.7524	65.7669	69.8975	71.1889	76.4303
PC	MSE	0.4675	0.6234	0.7324	0.7986	0.8994	0.9632	1.0417	1.1189	1.2069	1.2658
	MAE	32.2330	40.0678	45.2799	48.2811	52.9285	55.0638	57.8184	60.6103	63.6220	65.1578

Notes: To aid interpretation all values have been multiplied by 10^4. h represents the size of the step ahead returns forecast. Mean squared error (MSE) and mean absolute error (MAE) shown for 1–10 step ahead forecasts. For h = 1–10, each model is tested for equal forecast predictability against the random walk using the (i) Diebold–Mariano (1995) test (MSE and MAE) (ii) the Clark and West (2007) mean squared prediction error test. † denotes a failure to reject the null hypothesis at the 10% significance level for the Diebold–Mariano test. In all cases the Clark and West (2007) null hypothesis of equal forecast predictability is rejected in favour of the model at the 1% significance level.

Table 6
30-min Out-of-sample forecasts versus the random walk model – 12-month maturity.

Model	h	1	2	3	4	5	6	7	8	9	10
Random walk	MSE	1.1920	1.4910	1.6963	1.7757	1.9687	2.0332	2.1749	2.2697	2.4230	2.4676
	MAE	44.1535	55.4963	63.5706	66.3762	73.1473	75.1118	79.1637	82.0378	86.5897	87.1944
AR	MSE	0.3950	0.4427	0.5061	0.5603	0.6208	0.6697	0.7405	0.7789	0.8320	0.8764
	MAE	26.8225	32.1397	35.7122	38.0839	41.1539	43.0178	45.5267	47.4962	49.5999	50.9387
ARFIMA	MSE	0.3537	0.4295	0.4977	0.5499	0.6124	0.6612	0.7191	0.7710	0.8258	0.8701
	MAE	27.5947	33.2719	37.0554	39.5258	42.5802	44.4162	46.7676	48.6743	50.7631	52.0845
VAR	MSE	0.4087	0.4429	0.5115	0.5583	0.6241	0.6903	0.7284	0.7784	0.8327	0.8787
	MAE	27.4846	32.5762	36.0204	38.2088	41.1966	43.2801	45.5979	47.4904	49.6439	50.9674
VECM	MSE	0.3925	0.4396	0.5132	0.5612	0.6413	0.6719	0.7343	0.7848	0.3925	0.4396
	MAE	28.3534	33.8653	37.7077	40.2179	43.7459	45.8064	48.2579	50.1825	28.3534	33.8653
GARCH	MSE	0.3681	0.4564	0.5170	0.5642	0.6264	0.6771	0.7364	0.7897	0.8398	0.8845
	MAE	26.3708	32.1177	35.7612	38.3616	41.3713	43.4038	45.7914	47.9596	49.9710	51.4690
TGARCH	MSE	0.3653	0.4520	0.5185	0.5716	0.6294	0.6804	0.7394	0.7955	0.8502	0.8977
	MAE	26.4471	32.2984	36.0351	38.7155	41.7502	43.8371	46.2818	48.5314	50.6453	52.2164
VAR-BEKK	MSE	0.5066	0.5588	0.6543	0.6122	0.8142	0.8513	0.9228	1.0646	0.9105	1.0409
	MAE	30.1545	34.0260	39.3149	39.1157	46.4218	48.0023	50.8235	54.7364	52.4685	55.7775
PC	MSE	0.4130	0.5307	0.6102	0.6633	0.7283	0.7706	0.8283	0.8849	0.9452	0.9818
	MAE	26.3579	32.7688	37.0626	39.2906	43.0286	44.5462	47.0734	49.3016	51.8207	52.6622

Notes: To aid interpretation all values have been multiplied by 10^4. h represents the size of the step ahead returns forecast. Mean squared error (MSE) and mean absolute error (MAE) shown for 1–10 step ahead forecasts. For h = 1–10, each model is tested for equal forecast predictability against the random walk using the (i) Diebold–Mariano (1995) test (MSE and MAE) (ii) the Clark and West (2007) mean squared prediction error test. In all cases the null hypothesis of equal forecast predictability is rejected in favour of the model at the 1% significance level.

counterpart, with the 1-month adjusted R^2 noticeably higher than the next highest value.

5. Out-of-sample forecasting performance

5.1. 30-min Out-of-sample forecasting performance

Although some predictability is found in-sample, particularly at a 30-min frequency, we still need to investigate whether this extends to an out-of-sample context. Whilst in subsequent sections of this paper, the last year of data is reserved as the out-of-sample forecasting period; to utilise as much of this novel dataset as possible, and given the intraday in-sample period of the initial 2112 observations, for this section alone we treat the remaining 46,000 observations as out-of-sample. Employing the 30-min data, Tables 3–6 present the mean square error (MSE) and mean absolute error (MAE) of all models, forecasting h period changes in IV, where $h = 1,\ldots,10$; in other words, we are employing rolling forecasts from 30 min to 5 h ahead. The Diebold–Mariano (1995) test using MSE and MAE and the Clark and West (2007) mean square prediction error (MSPE) test are implemented to assess predictive ability for each forecasting model against the random walk for each period h. The former is applicable in a wide variety of settings including where forecast errors are non-Gaussian, non-zero-mean, or serially or contemporaneously correlated. However when one model nests the other this test tends to be poorly sized (see Clark and West, 2006). The latter test addresses this issue by adjusting the point estimates of the difference between the MSPEs of the two models for the noise associated with the larger of the two models. For both tests the null hypothesis of equal predictive ability is tested against the alternative hypothesis that the random walk is outperformed using standard normal critical values.

Table 7
Ranking by loss functions for 1-day ahead forecasts.

Model	1 month		3 months		6 months		12 months	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Random walk	21	21	21	21	21	21	21	21
30-min Models								
AR	18	17	19	14	19	9	18	13
ARFIMA	20	18	20	20	20	13	20	19
VAR	11	11	14	10	5	1	1	1
VECM	15	13	16	19	15	12	11	12
AR-GARCH	17	19	18	16	18	7	13	4
AR-TGARCH	16	20	17	17	16	5	15	2
VAR-BEKK	13	14	11	13	17	20	9	9
PC	14	15	12	8	13	15	16	14
Daily models								
AR	8	7	5	3	3	4	8	3
ARFIMA	5	6	4	5	9	14	10	15
VAR	12	10	3	6	4	10	14	18
VECM	19	16	13	18	14	19	19	20
AR-GARCH	9	5	8	9	8	8	4	5
AR-TGARCH	7	4	7	7	7	6	6	6
VAR-BEKK	10	12	15	15	11	18	7	10
PC	6	8	6	2	12	16	17	16
ECONOMIC	4	9	1	1	1	17	12	17
Half-daily models								
OLS Dummy	1	3	2	4	2	11	5	11
AR-GARCH	3	2	10	12	10	3	2	7
AR-TGARCH	2	1	9	11	6	2	3	8

Notes: Models ranked for each maturity for both mean squared error (MSE) and mean absolute error (MAE) from lowest loss (rank 1) to highest.

Strikingly in all cases the random walk benchmark is outperformed.[13] This result is perhaps unsurprising given our in-sample results uncovered the typical persistence found in volatility. That, out-of-sample, this persistence can be used over short-horizons, to enhance forecasts is useful information to intraday traders.

In terms of how the competing models perform, the TGARCH and PC models yield the lowest MSE and MAE respectively at the 1-month maturity for 1 and 2 period changes (Table 3). For longer period returns, the ARFIMA model and the VAR model perform better under MSE and MAE respectively. At the 3-month maturity (Table 4), the TGARCH and PC models again perform best for the 1-period change for MSE and MAE respectively, with the VAR model having the lowest MSE (MAE) for 6 from 9 (5 from 9) of the remaining returns. Moving to the 6-month maturity (Table 5), the TGARCH and PC models performance continue to dominate at the 1 period change with the VAR model performing well overall – the latter yielding the lowest MSE (MAE) for 5 from 9 (6 from 9) of the remaining 2 to 10 period returns. Finally, at the 12-month maturity (Table 6), under the MSE criteria the ARFIMA and VECM models yield the lowest loss with the ARFIMA dominating from 1 to 8 periods ahead whilst for the MAE the PC model again performs well for the 1 period change but AR, VAR and VECM models are preferred for longer maturity changes.

In sum, in each 30-min case, the random walk performs poorly yielding the highest sample loss versus alternative models. Across both loss functions at the shortest return horizon the tendency is for the ARCH-type models to perform well whilst predominantly the VAR and ARFIMA models are stronger between the 2–10 period returns.

5.2. Tests for superior predictive ability and the model confidence set

Given that the forecasting models are able to outperform the random walk at the 30-min level over short (within the day) horizons, we proceed to test whether modelling using intraday data or the related seasonality patterns produces superior one-day ahead forecasts to daily data. This is useful for the following reasons: (i) as a comparison with the typical forecasting exercise in the literature over a daily frequency; (ii) estimates of daily volatility are used in 10-day VaR calculations for regulatory purposes and (iii) assessing whether intraday information usefully augments forecasting models at the conventional daily frequency provides a new perspective on market efficiency issues.

To facilitate our prediction exercise we therefore forecast one-day ahead changes in IV from the daily (i.e., based on one observation ahead), the half-daily ON (i.e., based on 2 observations ahead) and the 30-min (i.e., based on 48 observations ahead) datasets. Subsequently, to allow an equal comparison of all various competing models and data frequencies, the out-of-sample forecasting period is now restricted to be the final year of available data (26 September 2007 to 25 September 2008) and we apply Hansen's (2005) SPA test.[14] This test allows for the controlling of the full set of models and their interdependence when evaluating the significance of relative forecasting performance. The null hypothesis is that the forecast under consideration (i.e., the benchmark) is not inferior to any alternative forecast. This test builds on the framework of White (2000) but instead uses a sample-dependent distribution under the null hypothesis. Hansen shows that this results in a more powerful test which is less sensitive to the inclusion of poor alternative models. As in the previous sections, we apply the MSE and MAE loss functions.

Table 7 ranks the models by the size for each loss function where the higher the rank the smaller the loss. Table 8 shows the SPA results for the 1 and 3-month maturities and Table 9, the 6 and 12-month maturities. Taking Table 7 first and therefore leaving inference aside for the moment, results show that the random walk model always presents the highest values for each loss function and is therefore outperformed by all other models. Top ranking at the 1-month maturity is always a half-daily model whilst for both loss functions at the 3-month maturity and the MSE at the 6-month maturity, it is the economic model that dominates. However, at the 6-month maturity using the MAE, and at the 12-month maturity for both loss functions, the 30-min VAR gives the lowest sample loss. Examining models ranked in the top three, half-daily models continue to perform particularly well at the 1-month maturity with these models ranked in the top three in 6 out of 6 cases, whilst daily models perform well at the 3-month maturity. For the 6 and 12-month horizon the picture is more mixed, with the top three positions being spread between half-daily, daily and the 30-min models.

Nevertheless, with the notable exception of the VAR model, the 30-min models do not typically perform particularly well across the maturities when ranked against the daily and half-daily ON models. This can be seen by comparing identical models from Table 7 and considering both loss functions simultaneously: daily models are ranked between 1st to 10th position more than three times as often as the 30-min models. Conversely, the 30-min models are ranked at below 10th position almost twice as often as daily models. Such poor performance might be *a priori* expected as the 30-min models are being asked to forecast usefully 48 periods ahead to cover one-day; an unlikely proposition for any short order parametric model.

[13] Individual test statistics are omitted to save space but can be provided on request. For the Diebold and Mariano (1995) test there are two cases (out of 640) where we are unable to reject the null of equal predictability both of which are overturned by the Clark and West (2007) test (see Table 5).

[14] For the 30-min dataset, the initial in-sample period employed as the initial input to this forecasting exercise are the 2112 observations prior to 26 September 2007 (i.e., 25/7/2007-25/9/2007).

Table 8
Test for SPA and MCS for 1-day ahead forecasts – 1 and 3-month maturity.

Model	1 months				3 months			
	MSE	SPA *p*-values	MAE	SPA *p*-values	MSE	SPA *p*-values	MAE	SPA *p*-values
Random walk	33.1979†	0.0000***	423.0125†	0.0000***	16.0818†	0.0000***	286.9061†	0.0000***
30-min Models								
AR	16.2463	0.0110**	290.7203	0.0030***	8.2079†	0.0050***	204.1255	0.0120**
ARFIMA	16.4576	0.0320**	291.2387	0.1820	8.5800†	0.0240**	206.6171†	0.0120**
VAR	15.8811	0.3440	286.4846	0.6410	8.0605	0.2610	201.4995	0.6340
VECM	16.0902	0.1920	287.4710	0.4790	8.1420	0.2960	205.8762	0.1900
AR-GARCH	16.2352	0.0970*	292.0051	0.0530*	8.1683†	0.0540*	204.8295†	0.0420**
AR-TGARCH	16.2116	0.1090	292.1753	0.0630*	8.1638	0.0520*	205.0458†	0.0330**
VAR-BEKK	16.0313	0.2060	288.3575	0.4900	7.9374	0.5860	202.0133	0.5750
PC	16.0451	0.1790	289.1294	0.2880	7.9478	0.5590	200.8276	0.7770
Daily models								
AR	15.7755	0.6460	284.1487	0.5400	7.8345	0.9640	199.5152	0.9600
ARFIMA	15.6959	0.5950	284.1410	0.7370	7.8048	0.8920	200.1547	0.8610
VAR	15.8924	0.3410	285.7945	0.5760	7.7882	0.8750	200.3229	0.7840
VECM	16.3934	0.0740*	290.3074	0.2500	8.0500	0.3410	205.7340	0.1600
AR-GARCH	15.7913	0.6250	283.3061	0.8880	7.9093	0.5890	200.8943	0.4530
AR-TGARCH	15.7704	0.6640	283.2348	0.9020	7.9040	0.5210	200.6212	0.4210
VAR-BEKK	15.8488	0.4450	287.3561	0.4210	8.1230	0.1410	204.7917	0.0520*
PC	15.7069	0.6500	284.9831	0.5880	7.8460	0.8430	199.0452	0.8350
ECONOMIC	15.5133	0.6330	285.1522	0.7620	7.7665	0.8330	199.0153	0.9830
Half-daily models								
OLS Dummy	15.1734	0.9710	281.2404	0.8170	7.7708	0.9660	199.8247	0.9150
AR-GARCH	15.3401	0.3010	280.6596	0.8750	7.9355	0.4950	201.8900	0.5990
AR-TGARCH	15.3100	0.6890	280.5553	0.9850	7.9305	0.4910	201.8168	0.5980

Notes: To aid interpretation all MSE and MAE values have been multiplied by 10^4.
* Rejection of the SPA null that the forecast under is not inferior to any alternative forecast at 10%.
** Rejection of the SPA null that the forecast under is not inferior to any alternative forecast at 5%.
*** Rejection of the SPA null that the forecast under is not inferior to any alternative forecast at 1%.
† Rejection from the MCS at the 10% confidence level with 10,000 bootstrap replications.

Table 9
Test for SPA and MCS for 1-day ahead forecasts – 6 and 12-month maturity.

Model	6 months				12 months			
	MSE	SPA *p*-values	MAE	SPA *p*-values	MSE	SPA *p*-values	MAE	SPA *p*-values
Random walk	12.5687†	0.0000***	266.0367†	0.0000***	9.0584†	0.0000***	212.1870†	0.0000***
30-min Models								
AR	6.1017	0.0710*	175.4749	0.0280**	4.3649	0.0100**	148.5348	0.0100**
ARFIMA	6.2696	0.1260	176.1451	0.3340	4.7208	0.0610*	150.9671	0.2140
VAR	5.9006	0.8930	172.1028	1.0000	4.2519	0.9830	146.1364	0.9960
VECM	6.0104	0.5520	176.0085	0.2760	4.3009	0.7950	148.4820	0.5380
AR-GARCH	6.0391	0.2570	175.1972	0.2360	4.3093	0.8010	147.1944	0.8420
AR-TGARCH	6.0197	0.3350	174.9303	0.2890	4.3231	0.6270	147.0559	0.8500
VAR-BEKK	6.0362	0.3800	178.6450	0.0670*	4.2829	0.8440	148.2075	0.4630
PC	6.0000	0.5890	176.7129	0.1110	4.3247	0.5050	148.5893	0.3320
Daily models								
AR	5.8887	0.9330	174.9115	0.7220	4.2770	0.9000	147.0887	0.8760
ARFIMA	5.9189	0.6380	176.5593	0.3770	4.2864	0.9260	149.0397	0.6480
VAR	5.8986	0.7710	175.8158	0.4980	4.3144	0.8440	149.3346	0.5880
VECM	6.0035	0.5020	178.6031	0.2030	4.3841	0.4520	151.1639	0.2490
AR-GARCH	5.9185	0.1360	175.2498	0.1670	4.2620	0.9950	147.2949	0.9490
AR-TGARCH	5.9126	0.7820	175.1307	0.6350	4.2667	0.9070	147.3181	0.9640
VAR-BEKK	5.9331	0.6980	178.4659	0.1270	4.2742	0.9080	148.3163	0.7140
PC	5.9444	0.9680	176.8489	0.2070	4.3441	0.7210	149.1002	0.5240
ECONOMIC	5.7900	0.7090	177.4761	0.2880	4.3090	0.5050	149.1045	0.5160
Half-daily models								
OLS Dummy	5.8868	0.8710	175.9882	0.4180	4.2635	0.9970	148.3747	0.7530
AR-GARCH	5.9196	0.3290	174.4389	0.7200	4.2539	0.9370	147.3667	0.8310
AR-TGARCH	5.9095	0.7900	174.3226	0.7810	4.2598	0.0440**	147.4664	0.1910

Notes: To aid interpretation all MSE and MAE values have been multiplied by 10^4.
* Rejection of the SPA null that the forecast under is not inferior to any alternative forecast at 10%.
** Rejection of the SPA null that the forecast under is not inferior to any alternative forecast at 5%.
*** Rejection of the SPA null that the forecast under is not inferior to any alternative forecast at 1%.
† Rejection from the MCS at the 10% confidence level with 10,000 bootstrap replications.

Table 10
Pesaran–Timmermann predictive failure test for 1-day ahead forecasts – 30-min, half-daily, and daily datasets.

Model	1 month	3 months	6 months	12 months
Random walk	0.4291	0.4559	0.3755	0.4444
30-min Models				
AR	0.4521	0.4598	0.5172	0.4713
ARFIMA	0.4866	0.4521	0.4981	0.4713
VAR	0.5211	0.5249	0.5326	0.4904
VECM	0.4828	0.4636	0.5057	0.5019
AR-GARCH	0.4904	0.4291	0.5441*	0.4751
AR-TGARCH	0.5019	0.4406	0.5287	0.5019
VAR-BEKK	0.4483	0.4751	0.4981	0.4981
PC	0.4598	0.4943	0.4444	0.4674
Daily models				
AR	0.5249	0.4406	0.4751	0.4904
ARFIMA	0.5249	0.4751	0.4215	0.4406
VAR	0.5364	0.5019	0.4866	0.5019
VECM	0.4751	0.4828	0.4789	0.4981
AR-GARCH	0.4943	0.4291	0.4636	0.4483
AR-TGARCH	0.5057	0.4176	0.4713	0.4521
VAR-BEKK	0.4636	0.4483	0.4521	0.4751
PC	0.5134	0.4789	0.4444	0.4636
ECONOMIC	0.5172	0.5057	0.4559	0.4866
Half-daily models				
OLS Dummy	0.5594**	0.4828	0.4368	0.4483
AR-GARCH	0.5479*	0.4789	0.4713	0.4981
AR-TGARCH	0.5364	0.4981	0.4713	0.5019

Notes: Table shows the proportion of correct predictions for each forecast model.
* Rejection of the Pesaran–Timmermann null hypothesis of the predictive failure at 10%.
** Rejection of the Pesaran–Timmermann null hypothesis of the predictive failure at 5%.
*** Rejection of the Pesaran–Timmermann null hypothesis of the predictive failure at the 1%.

Turning to the significance of relative forecasting performance, Tables 8 and 9 present the SPA test using each model from each frequency as the benchmark. Starting with the results for a random walk benchmark, both tables corroborate and extend the findings at the 30-min frequency, as we always reject the null that the forecast under consideration is not inferior to any alternative forecast. Next using the daily models as benchmark across both MSE and MAE, we were unable to reject the SPA null for any maturity at the 5% level of better whilst at the 10% level there only two rejections. In a similar vein the half-daily ON models only reject the SPA null once across all maturities. These results are in stark contrast to the SPA *p*-values generated when the 30-min models are the benchmark where there are 20 intraday model rejections. Overall it is again clear that the 30-min models are frequently not performing as well as other models using data sampled at a lower frequency.

Finally Tables 8 and 9 also show the results of Hansen et al.'s (2011) MCS procedure that produces the 'best' set of models with a given level of confidence.[15] As the authors' discuss the set is akin to a confidence interval for a parameter, and is attractive as it acknowledges the limitations of the data and permits more than one model to be best. Thus informative data will yield the best model, whilst less informative data means it is not possible to distinguish between competing models, resulting in a larger confidence set.

[15] The MCS procedure uses a sequential testing procedre to construct the confidence set that relies on test statistics that have non-standard asymptotic distributions that are estimated by bootstrap methods. As the authors' discuss the genesis of the sequential testing to determine the number of superior models is done in a similar fashion to the trace test for the number of cointegrating relationships in a VAR. In our application of the MCS procedure the confidence level is set to 10% and the number of bootstrap replications to 10,000.

Table 11
Pesaran–Timmermann predictive failure test for 1-day ahead forecasts: performance by day-of-the-week.

	1 month	3 months	6 months	12 months
Panel A: Half-daily ON AR-GARCH				
Monday	0.5962**	0.4808	0.4808	0.5000
Tuesday	0.3846	0.4615	0.4808	0.5385
Wednesday	0.5660	0.5283	0.3962	0.4528
Thursday	0.4808	0.4423	0.4808	0.4615
Friday	0.7115*	0.4808	0.5192	0.5385
Panel B: Daily AR-GARCH				
Monday	0.5385	0.4615	0.5192	0.5000
Tuesday	0.4423	0.3462	0.4615	0.5192
Wednesday	0.5283	0.4528	0.4151	0.4717
Thursday	0.4231	0.5385	0.4615	0.3846
Thursday	0.4231	0.5385	0.4615	0.3846
Friday	0.5385	0.3462	0.4615	0.3654

Notes: Table shows the proportion of correct predictions for the Half-Daily ON AR-GARCH (Panel A) and Daily AR-GARCH (Panel B) models for each day of the week. The day of the week refers to the point at which the forecast is formed. For example, the row for Monday shows the proportion of directionally correct next day forecasts based on information available at 4 pm Monday London time.
* Rejection of the Pesaran–Timmermann null hypothesis of the predictive failure at the 10%.
** Rejection of the Pesaran–Timmermann null hypothesis of the predictive failure at 5%.
*** Rejection of the Pesaran–Timmermann null hypothesis of the predictive failure at 1%.

The results of the MCS procedure indicate that the random walk never enters the confidence set, which given prior performance is not surprising. However for the remaining models we can see that it is only the intraday models that fail to enter the MCS, with all rejections occurring at the 3-month maturity. At this maturity, the MCS and SPA results are complimentary insofar as each time a model fails to enter the MCS, the same model is rejected under the SPA null. However, for the remaining maturities all models enter MCS and thus the MCS framework distinguishes less between models of different frequencies than the SPA alternative. Of course, MSE and MAE that are integral to both the SPA and MCS are purely statistical criteria, and sign type tests are perhaps more appropriate to assess trading success. Therefore to explore whether, in particular, half-daily ON and daily models can be delineated, such an analysis follows in the next section.

5.3. Tests for predictive failure

A logical step when attempting to forecast financial time-series is to implement tests of economic significance. We begin exploring this issue by examining the ability of the forecast models to detect daily changes in the direction of implied volatility. This is achieved via the predictive failure test of Pesaran and Timmermann (1992) that is based on the proportion of times that the sign of a series is correctly forecast. A rejection of the null hypothesis of predictive failure, using standard normal critical values, implies that the forecasted and actual change in IV are not independently distributed, in other words that the former can predict the direction of the latter. Table 10 presents the ratio of correct predictions, and the outcome of the predictive failure test.

We can see that all models predict the changes in IV in the range of 38–56%. Moving to the predictive failure test itself, we find no evidence that the random walk or daily models can aid in predicting the change in IV. Strikingly though, evidence is found at the 1-month horizon for the half-daily ON and ON AR-GARCH models, and the 6-month horizon for the GARCH 30-min model. In particular, the evidence of the half-daily ON models is interesting given the observed behaviour of IV in Figs. 2 and 3.

Table 12
Pseudo trading strategy.

Threshold filter	1 month	3 months	6 months	12 months
Panel A: Half-daily ON AR-GARCH				
0	0.1300	0.0233	0.0006	0.0576
0.005	0.1394	−0.0020	0.0064	0.0126
0.01	0.1159	−0.0065	−0.0063	−0.0082
Panel B: Daily AR-GARCH				
0	−0.0097	−0.0717	−0.0182	−0.0269
0.005	0.0225	−0.0052	−0.0028	−0.0025
0.01	0.0537	0.0000	0.0000	0.0000

Threshold filter	Half-Daily ON AR-GARCH	Daily AR-GARCH
Panel C: Friday forecast		
0	0.1113	−0.0174
0.005	0.1093	−0.0168
0.01	0.1021	0.0138

Notes: Panels A and B show the cumulative 'vols' generated by trading on forecasts of every weekday from the Half-Daily ON AR-GARCH and Daily AR-GARCH models respectively. The threshold filter denotes the absolute value of a forecast required to generate a trade. A larger filter value requires a larger forecasted change in IV to generate a trading signal. Panel C shows the cumulative 'vols' generated by 4 pm Friday to 4 pm Monday forecasts only.

To further illustrate the difference between the half-daily and daily models, it seems appropriate to explore predictability by day-of-the-week yielding a comparison of how, each day, a model can forecast 1 trading day ahead. Therefore Table 11 presents the results of the Pesaran and Timmermann test on the decomposed day-of-the-week forecasts for two representative models – the half-daily ON AR-GARCH and the daily AR-GARCH. Panel A gives the half-daily model and impressively Friday and Monday at the 1-month maturity show significant ratios of correct forecasts of 71% and 60% respectively; percentages much higher than found at the aggregate level and providing more evidence of a weekend effect. Of course, the possibility of a weekend effect was illustrated initially in Figs. 2 and 3. Across other maturities no significant results are found. This lack of significance is replicated in Panel B, where the tests for the daily model are shown.

5.4. Pseudo trading strategy

To further examine the economic significance of our forecasting models it is desirable to use these forecasts to implement a trading strategy. However, in our current context, it is not strictly possible as the construction of a full strategy would require more information on the IV surface.[16] Instead we can gain further insight as to the potential profitability of the overnight models by implementing a pseudo trading strategy that calculates the quantity of volatility points (i.e., volatility net profit or 'vols' as they are called by option traders; see Dunis and Huang, 2002) generated by acting upon our forecasts and ignoring any out-of-the-money data issues.

Although an imperfect approach, this nevertheless provides extra information above-and-beyond that of the Pesaran and Timmermann test. The 'vols' are calculated as follows: (i) A long (short) position is taken when a forecast predicts a positive (negative) change in IV of greater than a stated threshold (0%, 0.5%, and 1%); (ii) The next trading day we "close" the position using the observed ATM IV ignoring any strike price mismatch; (iii) For each trade the 'vols' gained or lost are recorded and summated across the forecast period.[17] Table 12 presents the results.

Panels A and B show the cumulative 'vols' based on forecasting every weekday in our out-of-sample period using our two representative models.[18] The implications are clear; both forecast models perform best at the 1-month maturity. Across all thresholds at this shorter maturity, the ON AR-GARCH model yields cumulative 'vols' in excess of 10% which contrasts with the daily AR-GARCH model with average cumulative 'vols' of 2.22%. These results confirm our previous results, again pointing to the superior performance of the ON models in forecasting changes in IV at the 1-month maturity. For maturities greater than 1 month, the cumulative 'vols' by both forecast models are modest at best, and occasionally negative.

Turning to Panel C we show the cumulative 'vols' based on 4 pm Friday to Monday forecasts alone. This allows us to examine the contribution of forecasts generated on a Friday to the cumulative 'vols' in Panels A and B. For example, using a 1% threshold, we can see in Panel C that the ON AR-GARCH model returns 10.21%. This compares with 11.59% shown in Panel A for the same model using the full week. Clearly, we find that for the ON model, a large proportion of cumulative 'vols' are generated from the 4 pm Friday to Monday forecasts, thus complimenting the findings of Tables 10 and 11.

Overall, the results of the pseudo strategy, and those in the previous sections, imply there is a clear role for higher (than daily) frequency data in forecasting the daily change in IV, at least at the 1-month horizon. In this context, the results are suggestive of a *type* of informational inefficiency; there is past diurnal information yet to be impounded into the 4 pm "daily" price!

6. Conclusions

Understanding the evolution of implied volatility is important as it has clear implications for the behaviour of asset prices and for practitioners in terms of profitable trading strategies. However, to our knowledge, the extant literature has not investigated the forecastability of high frequency, intraday implied volatility. Therefore we employ a dataset sampled at a 30-min frequency from the EUR–USD options market over the period November 2004 to September 2008. A number of maturities are covered and for each maturity over 48,000 observations are available. Moreover, a variety of forecasting models are applied to our 30-min dataset and two other comparator datasets derived from this high frequency: a half-daily sampled dataset allowing the assessment of intraday seasonalities such as overnight effects (ON) and a daily sampled dataset to represent the conventional frequency. In particular, the preliminary in-sample approach suggests London ON effects (4 pm–4 am) may usefully augment predictability regressions.

To begin we find evidence that the use of 30-min frequency data provides a useful tool for out-of-sample forecasting at shorter horizons, within a given day. The models applied at this stage include a number of time-series models, both homoscedastic and heteroskedastic; the latter allowing the explicit modelling of the

[16] For example, today's ATM option maybe out-of-the money tomorrow, hence to unwind the trade, IV on out-of-the money options data are also required. Of course, in any financial time series context, it is arguable whether reliable tests of economic significance are ever fully possible given issues such as transaction costs and price impact.

[17] Note that the strategy is closest in spirit to the Pesaran and Timmermann (1992) test when the threshold is set to 0. It complements this test by providing a magnitude (the realized change in IV) from acting on forecasted changes in direction of IV across both correct and incorrect forecasts, as opposed to the binary outcome of calculating the proportion of times a model forecasts the direction of IV correctly.

[18] In generating these cumulative 'vols' in Panels A and B we note that for the two non-zero thresholds the majority of trades are generated at the 1-month maturity. By contrast for longer maturities there are a very modest number of trades with the daily model failing to generate a buy or a sell order for the largest trading threshold. Perhaps this is unsurprising given (i) the magnitude of the weekend effect appears most pronounced at the 1-month maturity (ii) the standard deviation of the 1-month implied volatility is over 40% greater than the other maturities.

'volatility' of volatility. Strikingly, all models clearly outperformed the random walk benchmark when forecasting up to 5 h ahead.

Moving on, we proceed to test whether modelling using intraday data or the related seasonality patterns produces superior one-day ahead forecasts to daily data. Applying Hansen's (2005) test for superior predictive ability (SPA), 30-min models as might be *a priori* expected, are typically found to provide inferior forecasts when employed to forecast 48 periods ahead. By contrast, and as shown by Hansen et al.'s (2011) model confidence set (MCS) procedure, forecasts based on models from either the daily or half-daily frequency (i.e., augmented by London ON effects) are seldom outperformed based on standard loss functions.

To explore whether, in particular, ON and daily models can be delineated, we subsequently employ the predictive failure test of Pesaran and Timmermann (1992) and a pseudo trading strategy. In the former, where we examine the ability of models to predict the direction of implied volatility, no support is found for daily models, with the most support garnered from ON models at the 1-month maturity. Moreover, decomposing the models into days-of-the-week show a clear day-of-the-week and maturity effect, with the main finding that the models sampled at the half-daily frequency were able to predict the sign change over the weekend, whilst the daily models were not. The results of the pseudo trading strategy indicate that both ON and daily models perform better at the shorter 1-month maturity. However, the main result from the pseudo trading strategy is that the ON model dramatically outperforms the daily counterpart at this maturity, complementing the findings of predictive failure test.

The findings of this study have clear implications for both FX traders and policy makers. In the case of the former we demonstrate that in a given trading day, forecasts based intraday data can be useful. In the latter case, the performance of our forecast models are important given IV is used in VaR calculations for regulatory purposes. However, by far the most far reaching implication of our study stems from our finding that ON models outperform daily models in terms of likely trading success. This would interest both FX traders and policymakers, as it suggests the existence of meaningful information that is not being exploited at the daily frequency and thus indicates a type of market inefficiency.

Acknowledgements

We are grateful to Ana-Maria Fuertes, the participants of the Canadian Economics Association 2012 and Forecasting Financial Markets 2011/2012 conferences and, in particular, to an anonymous referee for very helpful comments resulting in a much improved paper. Any errors remain our own.

Appendix A. Summary of in-sample results: daily and intraday

Panel A1: Summary of daily in-sample results

AR	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant lags	1/2	0/1	0/1	1/3
Adj. R^2	0.0219	−0.0007	0.0000	0.0396
ARFIMA	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
d	−0.1971***	−0.1163***	−0.1152***	−0.141***
VAR	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant own lags	4/4	2/4	3/4	2/4
Significant other lags	4/12	2/12	5/12	2/12
Adj. R^2	0.0839	0.0387	0.0526	0.0838
VECM	$IV_{1,t}$	$IV_{3,t}$	$IV_{6,t}$	$IV_{12,t}$
Significant cointegration	2/3	0/3	1/3	0/3
Significant own lags	2/3	1/3	1/3	2/3
Significant other lags	3/9	1/9	2/9	3/9
Adj. R^2	0.0988	0.0646	0.0824	0.1071
AR-GARCH	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant lags	2/2	1/1	1/1	1/3
γ_1	0.1039***	0.1922***	0.1992***	0.2276***
γ_2	0.8472***	0.7809***	0.8091***	0.7926***
Adj. R^2	0.0201	−0.0037	−0.0034	0.0085
AR-TGARCH	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant lags	2/2	1/1	1/1	1/3
γ_1	0.1247***	0.2200***	0.2077***	0.2318***
γ_2	0.8484***	0.7812***	0.8094***	0.7938***
γ_3	−0.0646*	−0.0622	−0.0211	−0.0129
Adj. R^2	0.0205	−0.0032	−0.0032	0.0088
VAR-BEKK	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant own lags	0/1	1/1	1/1	1/1
Significant other lags	1/3	2/3	0/3	1/3

(*continued on next page*)

Appendix A. (*continued*)

AR	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant multivariate ARCH and GARCH coefficients: 7/8				
PC	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant PCs	0/2	0/2	0/2	1/2
Adj. R^2	−0.0003	−0.0014	0.0091	0.0167
ECONOMIC	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
ΔOIL_{t-1}	0.1255*	0.1076**	0.0663	0.0611
ΔIV^{+}_{t-1}	−0.1839	−0.1278	−0.1771	−0.3163*
ΔIV^{-}_{t-1}	0.0961	0.0917	0.1169	0.1916***
ΔS^{-}_{t-1}	−0.7103*	−0.5689**	−0.2813	−0.2553
Adj. R^2	0.0114	0.0010	0.0104	0.0430
Panel A2: Summary of intraday in-sample results				
AR	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant lags	5/9	5/5	5/8	5/5
Adj. R^2	0.3467	0.1138	0.4034	0.1680
ARFIMA	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
d	−0.0167	−0.1671***	0.0911	−0.1385***
VAR	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant own lags	4/4	4/4	4/4	4/4
Significant other lags	2/12	5/12	2/12	4/12
Adj. R^2	0.3358	0.1277	0.3792	0.1659
VECM	$IV_{1,t}$	$IV_{3,t}$	$IV_{6,t}$	$IV_{12,t}$
Significant cointegration	2/2	1/2	1/2	2/2
Significant own lags	3/3	2/3	3/3	3/3
Significant other lags	2/9	3/9	3/9	2/9
Adj. R^2	0.3385	0.1261	0.4406	0.1701
AR-GARCH	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant lags	9/9	4/5	7/8	4/5
γ_1	2.8083***	0.0795***	3.3087***	0.0501***
γ_2	−0.0170***	0.4817***	−0.0307***	−0.0665
Adj. R^2	−0.2842	0.0945	−0.1259	0.1130
AR-TGARCH	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant lags				
γ_1	0.1485***	0.0790***	9.5269***	0.0874***
γ_2	−0.0040	0.4816***	0.0165**	−0.0717
γ_3	5.6419***	0.0012	−9.3609***	−0.0512**
Adj. R^2	0.1270	0.0945	0.2268	0.1054
VAR-BEKK	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant own lags	1/1	1/1	1/1	1/1
Significant other lags	1/3	1/3	1/3	0/3
Significant multivariate ARCH and GARCH coefficients: 6/8				
PC	$\Delta IV_{1,t}$	$\Delta IV_{3,t}$	$\Delta IV_{6,t}$	$\Delta IV_{12,t}$
Significant PCs	1/2	1/2	1/2	1/2
Adj. R^2	0.1758	0.0812	0.1138	0.1018

Notes: Full results available upon request. "Significant lags" yields the proportion of significant lags at the 10% level of the number of lags selected via SIC. For the VECM model the "significant cointegration" yields the proportion of significant cointegrating relationships with the denominator determined via the Johansen testing procedure. For the economic model coefficients are shown where there is at least one significant result for any maturity.

* Rejection of zero null at 10%.

** Rejection of zero null at 5%.

*** Rejection of zero null at 1%..

References

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115–158.
Andreou, P., Charalambous, C., Martzoukos, S., 2010. Generalized parameter functions for option pricing. Journal of Banking and Finance 34, 633–646.
Bank of International Settlement, 2012. Statistical Annex. BIS Quarterly Review (December 2012).
Becker, R., Clements, A., White, S., 2007. Does implied volatility provide any information beyond that captured in model-based forecasts? Journal of Banking and Finance 31, 2535–2549.
Becker, R., Clements, A., McClelland, A., 2009. The jump component of S&P 500 volatility and the VIX index. Journal of Banking and Finance 33, 1033–1038.
Busch, T., Christensen, B., Nielsen, M., 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. Journal of Econometrics 160, 48–57.
Caporale, G., Gil-Alana, L., 2010a. Fractional integration and data frequency. Journal of Statistical Computation and Simulation 80, 121–132.
Caporale, G., Gil-Alana, L., 2010b. Long Memory and Fractional Integration in High Frequency Financial Time Series. Working Paper No. 10-10. Brunel University.
Chalamandaris, G., Tsekrekos, A., 2010. Predictable dynamics in implied volatility surfaces from OTC currency options. Journal of Banking and Finance 34, 1175–1188.
Christensen, B., Hansen, C., 2002. New evidence on the implied-volatility relation. European Journal of Finance 8, 187–205.
Clark, T., West, K., 2006. Using out-of-sample mean square prediction errors to test the martingale difference hypothesis. Journal of Econometrics 135, 155–186.
Clark, T., West, K., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.
Corrado, C., Miller, T., 2006. Estimating expected excess returns using historical and option-implied volatility. Journal of Financial Research 29, 95–112.
Covrig, V., Low, B.S., 2003. The quality of volatility traded on the over-the-counter currency market: a multiple horizons study. Journal of Futures Markets 23, 261–285.
Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.
Doyle, J., Chen, C., 2009. The wandering weekday effect in major stock markets. Journal of Banking and Finance 33, 1388–1399.
Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. Journal of Finance 53, 2016–2059.
Dunis, C., Huang, X., 2002. Forecasting and trading currency volatility: an application of recurrent neural regression and model combination. Journal of Forecasting 21, 317–354.
Engle, R., Kroner, K., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.
Engle, R., Sokalska, M., 2012. Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. Journal of Financial Econometrics 10, 54–83.
Ferraro, D., Rogoff, K., Rossi, B., 2012. Can Oil Prices Forecast Exchange Rates? NBER Working Paper 17998.
Gonçalves, S., Guidolin, M., 2006. Predictable dynamics in the S&P 500 index options implied volatility surface. Journal of Business 79, 1591–1635.
Granger, C., Joyeux, R., 1980. An introduction to long-memory models and fractional differencing. Journal of Time Series Analysis 1, 15–39.
Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.
Harvey, C., Whaley, R., 1992. Market volatility prediction and the efficiency of the S&P 100 index options market. Journal of Financial Economics 31, 43–73.
Hibbert, A.M., Daigler, R., Dupoyet, B., 2008. A behavioural explanation for the negative asymmetric return-volatility relation. Journal of Banking and Finance 10, 2254–2266.
Hosking, J., 1981. Fractional differencing. Biometrika 68, 165–176.
Kellard, N., Dunis, C., Sarantis, N., 2010. Foreign exchange, fractional cointegration and the implied-realized volatility relation. Journal of Banking and Finance 34, 882–891.
Kim, M., Kim, M., 2003. Implied volatility dynamics in the foreign exchange markets. Journal of International Money and Finance 22, 511–528.
Konstantinidi, E., Skiadopoulos, G., 2011. Are VIX futures prices predictable? An empirical investigation. International Journal of Forecasting 27, 543–560.
Konstantinidi, E., Skiadopoulos, G., Tzagkaraki, E., 2008. Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. Journal of Banking and Finance 32, 2401–2411.
Locke, P., Sayers, C.L., 1993. Intra-day futures price volatility: information effects and variance persistence. Journal of Applied Econometrics 8, 15–30.
Muzzioli, S., 2010. Option-based forecasts of volatility: an empirical study in the DAX-index options market. The European Journal of Finance 16, 561–586.
Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.
Pesaran, M., Timmermann, A., 1992. A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics 10, 461–465.
Pojarliev, M., Levich, R.M., 2008. Do professional currency managers beat the benchmark? Financial Analysts Journal 64, 18–33.
Poon, S.-H., Granger, C., 2003. Forecasting volatility in financial markets: a review. Journal of Economic Literature 41, 478–539.
Taylor, S., Pradeep, Y., Zhang, Y., 2010. The information content of implied volatilities and model-free volatility expectations: evidence from options written on individual stocks. Journal of Banking and Finance 34, 871–881.
White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.