AN ANALYSIS OF THE PERFORMANCE OF GENETIC PROGRAMMING FOR REALISED VOLATILITY FORECASTING

Zheng Yin, Conall O'Sullivan and Anthony Brabazon

School of Business, University College Dublin, Ireland

Abstract

Traditionally, the volatility of daily returns in financial markets is modeled autoregressively using a time-series of lagged information. These autoregressive models exploit stylised empirical properties of volatility such as strong persistence, mean reversion and asymmetric dependence on lagged returns. While these methods can produce good forecasts, the approach is in essence atheoretical as it provides no insight into the nature of the causal factors and how they affect volatility. Many plausible explanatory variables relating market conditions and volatility have been identified in various studies but despite the volume of research, we lack a clear theoretical framework that links these factors together. This setting of a theory-weak environment suggests a useful role for powerful model induction methodologies such as Genetic Programming (GP). This study forecasts one-day ahead realised volatility (RV) using a GP methodology that incorporates information on market conditions including trading volume, number of transactions, bid-ask spread, average trading duration (waiting time between trades) and implied volatility. The forecasting performance from the evolved GP models is found to be significantly better than those numbers of benchmark forecasting models drawn from the finance literature, namely, the heterogeneous autoregressive (HAR) model, the generalized autoregressive conditional heteroscedasticity (GARCH) model, and a stepwise linear regression model (SR). Given the practical importance of improved forecasting performance for realised volatility this result is of significance for practitioners in financial markets.

Keywords: Realised Volatility, Genetic Programming, High Frequency Data

1 Introduction

Volatility is an important concept in finance and has different implications depending on the perspective of the user. From an investment perspective, volatility is a measure of the degree to which returns tend to fluctuate. Traders would like to capture the volatility caused by positive returns, whereas in contrast, risk management is more concerned about the volatility caused by negative returns. Volatility is a key element in the pricing of derivatives, is a key input in determining portfolio weights in a portfolio optimisation model and is also a key input in the calculation of regulatory capital requirements under the Basel II accords.[1] Hence, many stakeholders have an interest in being able to model and predict volatility.

In a conventional volatility model, volatility is a latent variable that is often estimated parametrically from historical daily returns using discrete time GARCH models [1] or continuous time Stochastic Volatility models [2]. The term realised volatility can be broadly defined as the sum of in-

---

[1]Basel II are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision.

traday squared returns, measured at short intervals [3]. Such a non-parametric volatility estimator has been shown to provide an accurate estimate of the latent process that defines volatility [5] and therefore, through realised volatility estimation, the latent volatility process is theoretically observable from historical intraday returns.

Genetic Programming (GP) is a powerful model induction methodology which has been widely applied for symbolic regression [6, 7]. A number of studies have previously applied GP for volatility modeling [9, 11, 12, 13, 15] but there are still some important questions which have not been addressed.

In particular, market conditions have been documented in the finance literature as having a high correlation with volatility in a variety of settings. A sample of these studies include [16] which examined the relationship between trading volume and volatility, [17] which examined the relationship between the number of transactions and volatility, [18] examined the relationship between price range and volatility, [39] examined the relationship between interest rates and volatility, [20] examined the relationship between implied volatility and volatility and [21] which examined the relationship between the bid-ask spread and volatility. This study extends previous works by identifying a range of metrics on market conditions and allowing GP to use these as inputs in modeling volatility. The calculated realised volatility is modeled directly using GP and the one-day-ahead RV is forecasted. Forecasting results from GP are compared with those from benchmark models drawn from the finance literature.

1.1   Structure of Paper

The remainder of this contribution is organised as follows. Section 2 provides some background on volatility modeling and provides the motivation for applying the GP methodology to RV forecasting. Section 3 describes the data used in this study. The forecasting results are provided in Section 4 and finally, conclusions and opportunities for future work are discussed in Section 5.

2   Overview of Volatility Modelling

In this Section we overview three key items. Initially, we provide an introduction to the concept of

realised volatility. Then we briefly introduce current state-of-the-art approaches for the forecasting of realised volatility. Finally, we provide the motivation for the GP methodology adopted in this study.

2.1   Realised Volatility

Under the concept of RV, returns are assumed to be generated by the stochastic differential equation (Equation 1), which is a continuous-time stochastic process over a given time period. The time period is divided into i equally-spaced adjacent intervals and the quadratic variation is defined as the limit of the sum of squared returns over these intervals, as the length of the sampling intervals goes to zero, where t_i and t_{i-1} are adjacent intervals (Equation 2). This limit is well-defined in the case of the logarithm price process $p(t)$, which is a semi-martingale. In the general semi-martingale case, assuming some (mild) restrictions on the types of leverage, the quadratic variation is an unbiased estimator of the integrated variance, $\int_0^T \sigma^2(t)dt$, and the square root of the quadratic variation is called realised volatility.

$$dp(t) = \sigma(t)dW(t) \qquad (1)$$

$$\lim_{i \to \infty} \left(\sum_i (p(t_i) - p(t_{i-1}))^2 \right) \qquad (2)$$

Realised volatility can be used to measure the interdaily volatility by summing up the intraday squared returns at short intervals, such as five or fifteen-minute intervals [24]. This concept is very important to volatility modeling. It has been pointed out in [23] that the standard volatility models used for forecasting at the daily level cannot readily accommodate the information in intraday data. The models specified directly for intraday data generally fail to capture the longer interdaily volatility movements sufficiently well. In contrast, using RV allows us to model volatility using relatively high frequency data, and also permits capture of stylised facts concerning interday volatility [23, 24].

In an ideal world, the quadratic variation from shorter intervals (as per Equation 2) is always closer to the integrated volatility than the one calculated using longer intervals. However, returns measured

at intervals shorter than five minutes are plagued by spurious serial correlation caused by various market microstructure effects including asynchronous trading, discrete price observations, and the bid-ask bounce [5].

There are different sampling schemes to estimate the realised volatility as reviewed in [22]. In this study, the RV estimation approach in [25] is followed as we use the same futures index data, FTSE 100 prices. It is also noted that the RV estimated using this method [25] successfully captured the stylised long-memory effect inherent in volatility.

2.2 Conventional RV Forecasting Models

It is well documented in the finance literature that realised volatility is a highly persistent process which has a long memory. Conventional methods used in modeling RV include ARFIMA (Autoregressive Fractionally Integrated Moving Average) [23, 26], HAR (Heterogeneous Autoregressive) proposed by [27], the simple AR (Autoregressive) type model [28, 29], and SV (Stochastic Volatility) with volatility treated as observable [29]. Recently there have also been HAR-type extended models including the HAR-GARCH model proposed by [30], and HAR with a jump process as proposed by [31].

A broad series of empirical work [29, 30, 32] has sought to compare the performance of various RV forecasting models.

In [30], ARFIMA, HAR and HAR-GARCH are compared based on tick-by-tick transaction prices from S&P 500 index futures data (1985-2004) with HAR-GARCH producing the best forecasting performance in terms of several metrics including R^2, RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and RMSPE (Root Mean Squared Percentage Error). In [32], AR, ARFIMA and HAR are compared and HAR gives the best result in terms of RMSE, MAE and R^2. This conclusion is drawn on a dataset consisting of tick-by-tick series for USDCHF (1989 to 2003), S&P 500 Futures (1990-2007) and 30-year US Treasury Bond Futures (1990-2003). In [29], simple AR, SV and HAR are compared and HAR gives the best forecasting performance in terms of RMSE, MAE and other measures on a dataset of equity market indices of SPX and DJIA (1997-2011) and two exchange

rates CADUSD and USDGBP (1998-2011). The ARFIMA has been reported in [30] and [32] to give a similar performance as HAR, however, its estimation procedure is more complex.

2.3 Motivation for Applying GP to RV Modelling

RV transfers intraday return information to an observable volatility, and therefore allows volatility to be modeled directly. While traditional methods of RV modeling rely solely on lagged values of RV (see Section 2.2), it has been documented that trading volume, number of transactions, price range (including the range of open and close, high and low), bid-ask spread and implied volatility have predicative information / explanatory power for volatility.

It has been noted in [33] that different market information is likely to capture distinct subtle aspects of the volatility process, the relative prominence of which may vary over time. Also different market information may suffer to greater or lesser extents from market microstructure biases. A study by [33] indicates that using a combination of the outputs from a series of GARCH models, with different volatility predictors, could reduce the forecast errors in a range of examined stocks.

In prior works, most studies [34, 35, 36, 37, 38, 39, 33, 40] used market information to explain / forecast conditional volatility in a GARCH type framework. The market information was added in the conditional variance equation as an explanatory factor but the underlying model was linear. The nonlinear Granger causality test conducted in [41] shows there is extensive evidence of bidirectional feedback between volume and volatility which such approaches cannot capture.

In summary, while we have some knowledge of the likely set of explanatory variables (based on market conditions) from prior literature, we still lack a clear theoretical framework as to which of these variables are most important and how they should link together to form a quality model for forecasting of RV. This setting of a theory-weak environment suggests a useful role for powerful model induction methodologies such as GP [6, 8].

In this study, GP is used to select from a set of plausible explanatory variables as identified in the finance literature, and then link them to RV by si-

multaneously evolving a suitable functional form. The functional form returned from training is then used to forecast a one-day-ahead RV. The model is re-trained each day using most recent information as no assumption is made in the modeling process that the relative importance of each explanatory variable remains unchanged over time. The performance of the GP models is compared with a HAR model which only uses RV lagged information as inputs, a GARCH model, which models RV and its volatility together and a linear regression model, which uses the same explanatory factors that are used to train the GP model. It should be noted that given the importance of volatility forecasts across a range of investment decisions, even small improvements in forecast accuracy can have significant practical implications.

3 Data and Methodology

3.1 Background

The dataset used in this paper consists of the complete records for all quotes and trades of European-style FTSE 100 index option contracts and FTSE 100 index futures contracts in 2004 from Euronext-Liffe. The London International Financial Futures and Options Exchange (Liffe) was established in 1982 and was taken over by Euronext in January 2002 to form a market called Euronext-Liffe. Since 2000, all trading in financial contracts on Liffe takes place on an electronic limit order book system, called the Liffe Connect platform.

The datasets used in this study are large, 'ultra high frequency data' [10], consisting of 75,755,106 records in the case of the index option dataset (41,794,081 records relating to call options and 33,961,025 records relating to put options). The index futures dataset consists of 26,271,084 observations. All of this data is time stamped.

The futures traded price data was used for RV estimation and both the trade and quote information was used to calculate intraday metrics including trading volume, bid-ask information, price range and the number of transactions. FTSE 100 index options data is used for the implied volatility calculation. Interest rate information, specifically, LIBOR rates (overnight, one-week and six-month) for 2004, were collected from Datastream.

The estimated RV is illustrated in Figure 1 (refer to Section 3.1.1 for estimation details for this data). The first six months of the data is used for initial in-sample training with the out-of-sample testing taking place during the final six months (129 trading days) of the year. Each day's forecast of RV is determined using all data available up to and including the previous day. For the first day's out-of-sample forecast (commencing on the first day of July), data from January 9th to June 30th is used. For the last day's forecast (the last day of December), data from January 9th to December 30th is used. The first five trading days in January are excluded as lagged information is required in the modeling process.
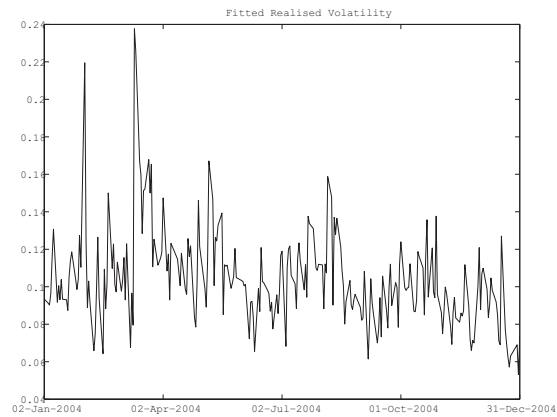

Figure 1: Annualised Daily Realised Volatility

3.1.1 Realised Volatility Estimation

FTSE 100 index futures traded from 8:00 am to 5:30 pm in 2004 and therefore there are 114 five-minute intraday returns each day in our dataset, which are calculated from the latest prices before each five-minute mark in Equation 3, where $\ln(p_{t,j})$ is the log price for the $j-th$ five-minute interval on day t and $r_{t,j}$ is the $j-th$ intraday return on day t with $j = 1, 2 \ldots, 114$.

$$r_{t,j} = \ln(p_{t,j}) - \ln(p_{t,j-1}) \qquad (3)$$

Let $r_{t,j}$, $0 \leq j \leq n$, represent a set of $n+1$ intraday returns for day t, so that $j = 0$ represents the closed-market period from the close on day $t-1$ until the open on day t, $j = 1$ represents the first five-minute on day t,... concluding with $j = n$ representing the final five-minute period on day t.

The realised volatility is used to measure the intraday return volatility. The realised variance for

trading day t, from the close on day $t - 1$ to the close on day t, is estimated by weighting the intraday squared returns as in Equation 4.

$$\sigma_t^2 = \sum_{j=0}^{n} \omega_j r_{t,j}^2 \qquad (4)$$

To ensure conditionally unbiased estimates when intraday returns are uncorrelated, so that $E[\sigma_t^2|\sigma_t^2] = \sigma_t^2$, it is necessary to apply the constraint $\sum_{j=0}^{n} \lambda_j \omega_j = 1$ [25]. The λ_j is the proportion of a trading day's total return variance that is attributed to period j and is assumed to be the same for all days t in the sample. To satisfy this constraint, ω is estimated as in Equation 5 and λ is estimated as in Equation 6.

$$\omega_j = \begin{cases} \frac{1}{(1-\lambda_0)n\hat{k}_j}, 1 \le j \le n \\ 0, j = 0 \end{cases} \qquad (5)$$

$$\begin{aligned} \hat{\lambda}_j &= \frac{\sum_t r_{t,j}^2}{\sum_t \sum_{i=0}^{n} r_{t,i}^2} \\ \hat{k}_j &= \frac{\sum_t r_{t,j}^2}{\sum_t \sum_{i=1}^{n} r_{t,i}^2} \end{aligned} \qquad (6)$$

The average annualised volatility from the estimated RV (in Figure 1.) is 10.37 percent, which is very close to the annualised daily return standard deviation, 10.26 percent, therefore, any potential bias caused by autocorrelation among intraday returns is small. The distribution of $\ln(\sigma)$ is almost symmetric and approximately Gaussian. Applying the augmented Dickey-Fuller test indicated that the realised volatility process does not contain a unit root as expected given the high degree of mean reversion in volatility. The low decline in the autocorrelations of the realised volatility series suggests a long memory process, another well documented stylised volatility fact.

3.2 Predictive Market Information

The relation between volatility and other exogenous market information has received increasing attention from academic researchers and a selection of the most commonly proposed explanatory variables are summarised in Table 1.

Based on this, the potential explanatory variables related to market conditions used in this study include price range, bid-ask spread, trading volume, number of transactions, corresponding implied volatility and interest rates. The lagged information in Table 2 is also used in the RV modeling process.

3.2.1 Implied Volatility

The implied volatility of at-the-money (ATM) one-month options is calculated from a volatility surface, which is fitted using all available trading information for FTSE 100 Index Options on the previous day. It is shown in Figure 2., and exhibits similar patterns as RV including long memory and persistence. However it should be noted that the implied volatility measure is an annualised estimate of the square root of the integrated variance over the subsequent one-month horizon and is significantly higher than the realised volatility estimates which is a single day measure of volatility even though this realised volatility value is also annualised. The difference between the implied volatility and the realised volatility is often referred to as the volatility risk premium. Thus the one-month ATM implied volatility may contain important information on the future path of the volatility so is a potentially useful explanatory variable to include in the modeling of realised volatility. Therefore lags of one to five periods of the implied volatility are included as explanatory variables in the modeling process.

3.2.2 Price Range

There are two daily information variables created for the price range category including the absolute difference of open price and close price and the absolute difference of the days highest price and the days lowest price. The squared daily return is also used as an explanatory variable. It is a proxy for the absolute price range from the close of adjacent days. The price range information is shown in Figure 5 and Figure 6. The squared daily returns are illustrated in Figure 7. These figures display the typical volatility clustering patterns often observed in financial time series with periods of market turbulence inter-dispersed with periods of tranquillity.
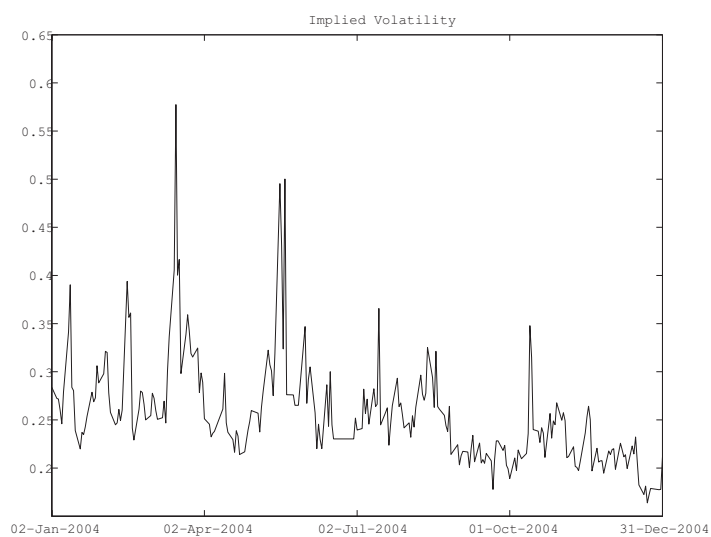
3.2.3 Trading Volume

There are two variables created for trading volume using the available data, namely total trading volume per day and average five-minute trading volume. The number of daily total transactions and the average trading duration (waiting time between

Table 1: Market/Economic Condition Variables Used in Volatility Forecasting

Volatility Forecasting Study	Predictive Variables Considered
[43]	Daily high-low range
[33]	Realised range, realised power variation
	Realised bipower variant and volume
[44]	Trading volume and implied volatility
[45]	5 Categories (38 variables) tested
[50] [49] [48] [47] [20] [46]	Implied volatility
[41]	Volume

This table gives market/economic information variables used in volatility forecasting in different studies besides the lagged volatility and return information.

Figure 2: Implied Volatility of ATM Option Expiry in One Month

trades) are also used. These variables are illustrated in Figures 8 to 11. It is clear that the spikes in trading volume tend to coincide with higher volatility periods as shown in the daily price range and squared return figures, although the trading volume/transaction figures show that large jumps decay away more slowly than the daily price range and squared return values. Furthermore when volatility is high the average trading duration is low as the number of trades tend to increase in these higher volatility periods.

3.2.4   Bid-Ask Spread

For bid-ask spread metrics, two variables are created using the available data, namely the daily average bid-ask spread, and the maximum bid-ask spread. These are illustrated in Figure 12. and Figure 13. As expected the maximum bid-ask spread is more volatile than the average bid-ask spread varying between 0.2 and 0.13 index points[1] whereas the average bid-ask spread is approximately 0.06 of an index point. It should be noted that the average index value over this period was approximately 4,523 thus the average bid-ask spread on the futures contracts in a proportionate sense is approximately 0.13 basis points (where one basis point is one hundredth of one percent) emphasising the high liquidity and low trading costs of FTSE 100 futures contracts.

3.2.5   Other Explanatory Variables

The squared RV is included as explanatory variable as squared RV is related to the RV's volatility, that is, the volatility of volatility itself. Overnight Libor, one week and six-month Libor are used as the nominal interest rate proxies and are illustrated in Figures 14-16.

Comparing all the figures, the implied volatility series (in Figure 2), the absolute difference of the daily highest and lowest price (in Figure 6) and daily total transaction number (in Figure 10) seem to be highly associated with the annualised RV (in Figure 1). However, these market variables are all related one to another. In general, when the volatility is high, the price range tends to be high, the transaction volume tends to be high, the trading duration tends to be small, the bid-ask spread tends to

be small and the interest rate tends to be low. The relationships invert when volatility is low.

3.3   GP Approach

In this study we employ GP for symbolic regression. The target variable is realised volatility and the evaluation of an individual GP tree is therefore an RV forecast. The available GP terminal set is outlined in Table 2 and the available function set is described in Table 3. Potential included variables in the GP trees include, five lagged values of RV, average daily trading (duration how often a trade occurs in a day), and Implied Volatility (IV) (estimated from the FTSE 100 index options data). The factors in Table 2 are from the previous day's information if no explicit lag indication given.

Each GP individual has a fitness value which indicates how well it performs when tested in-sample on the training dataset. The fitness function in this application is the mean squared error as defined in Equation 7, where RV_{target} is the target RV value, RV_{ind} is the evaluation of the individual, and $Number_{Days}$ is the number of data points in the training dataset.

$$Fitness = \sqrt{\frac{\sum(RV_{target} - RV_{ind})^2}{Number_{Days}}} \qquad (7)$$

In the experiments, all results are reported averaged across 10 runs and each GP run consists of 50,000 individuals evolved over 50 generations. The operation of the GP system is illustrated in Figure 3. In order to reduce the chance of over-fitting, the maximum tree depth is set to six, based on initial experimentation. The training process is summarized in Figure 4.

3.4   Benchmark Models

Below we briefly outline the three benchmark models against which we compare the evolved GP models. The models are drawn from extant studies which forecast RV and include the Heterogeneous Autoregressive model (HAR), a generalised autoregressive conditional heteroscedastic (GARCH) model, and a stepwise regression (SR) model.

---

[1]Each futures contract is worth 10 times the index when the futures contract expires so a bid-ask spread of 1 point in the futures contract is equivalent to a bid-ask spread of 0.1 point in the underlying index.

**Table 2**: Potential Explanatory Factors Used in GP

RV Lagged Information (one to five days lag)
Absolute Difference of Day Open and Close Price
Absolute Difference of Day Highest and Lowest Price
Daily Total Trading Volume
Average Five-minute Trading Volume
Daily Number of Transactions
Average Daily Absolute Difference of Bid and Ask Price
Maximum Daily Absolute Difference of Bid and Ask Price
Implied Volatility(IV) of a 1 Month at the Money Option
Average Daily Trading Duration in Seconds
Squared Daily Return
Squared RV
IV Lagged Information (two to five days lag)
Daily Libor
Weekly Libor
Six-month Libor
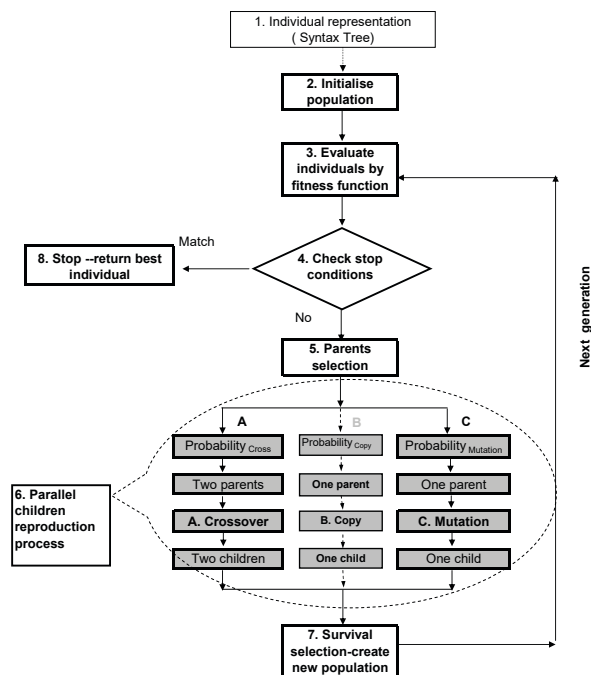
Figure 3: Flow Chart of GP Process
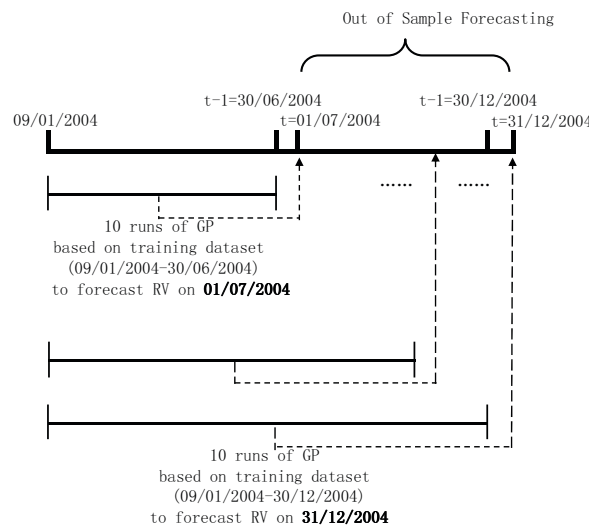
Figure 4: Overview of GP application for RV Forecasting

3.4.1 HAR model

In the Heterogeneous Autoregressive model (HAR) [32], RV is modeled using lagged information, including RV one day before, average RV in the last week and average RV in the last month. This model is provided in Equation 8, where c, α, β and γ are constant coefficients.

$$
\begin{aligned}
RV_t &= c + \alpha RV_{t-1} + \beta RV_w + \gamma RV_m \\
RV_w &= \tfrac{1}{5}\sum_{i=1}^{5} RV_{t-i}, \\
RV_m &= \tfrac{1}{21}\sum_{i=1}^{21} RV_{t-i}
\end{aligned}
\tag{8}
$$

The model coefficients are re-trained for each forecasted day.

3.4.2 GARCH Model

In the generalised autoregressive conditional heteroscedastic (GARCH) model, RV and its volatility (υ) are modeled jointly in Eqs. 9, in which Equation 9a is the mean equation and Equation 9c is the conditional variance equation. z_t in Equation 9b is a white noise process. c, α_i, β_i, k, ϕ_i and η_i are constant coefficients, and ε_{t-i} are independent identically distributed random variables sampled from a standard normal distribution.

$$
RV_t = c + \sum_{i=1}^{p} \alpha_i RV_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \nu_t
\tag{9a}
$$

$$
\nu_t = \upsilon_t z_t
\tag{9b}
$$

$$
\upsilon_t^2 = k + \sum_{i=1}^{p} \phi_i \upsilon_{t-i}^2 + \sum_{i=1}^{q} \eta_i \varepsilon_{t-i}^2
\tag{9c}
$$

3.5 SR Model

Stepwise regression (SR) is a systematic method for adding and removing potential explanatory variables from a multilinear model based on their statistical significance. The method begins with an initial model and then compares the explanatory power of incrementally larger and smaller models. At each step, the p value of an F-statistic is computed to test models with and without a potential factor. If a factor is not currently in the model, the null hypothesis is that the factor would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, the factor is added to the model. Conversely, if a factor is currently in the model, the null hypothesis is that the factor has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the factor is removed from the model.

In this stepwise regression model, all market information variables in Table 2 are considered as potential explanatory variables. The stepwise regression model is fitted for each day's RV forecast based on the in-sample training dataset, which is from the starting day until one day before the day to be forecasted and this is the same as GP model's training dataset as explained in Section 3.3. As shown in Equation 10 the factors in each day's forecasting model may vary in the stepwise selected linear regression model. By default there is always a constant intercept c.

Table 3: Function Set Available to GP

Addition
Subtraction
Multiplication
Division
Cumulative Distribution Function of Normal Distribution
Exponential Function
Nature Logarithm Function
Square Root
Cube Root
Sine Function
Cosine Function

$$RV_t = c + \alpha_1 Var_1 + \alpha_2 Var_2 \cdots + \alpha_i Var_i \quad (10)$$

4 Results

The out-of-sample results from our experiments are reported in this Section. Initially, we report the forecast errors for each modeling methodology, then we present a statistical analysis of these results. Finally, we report the results from a series of information encompassing tests.

4.1 Forecast Errors

The forecast errors are presented in Table 4. In this table, three measures of forecast error (MAE, MAPE and RMSE) are presented for GP and the benchmark models. The final column in the table presents the R^2 from the linear regression which regresses the actual RV against the predicted values from each method.

The results indicate that using the average of the GP models' predictions produces a smaller MAE, MAPE and RMSE and also a notably higher correlation in terms of R^2 than is the case for the three benchmark models.

Comparing the three benchmark models amongst themselves, the SR model performs slightly better than the HAR model and the GARCH model in terms of MAE and MAPE error metrics. Using RMSE as the error metric, we note that the HAR model performs slightly better than the GARCH model and SR models. Considering R^2, HAR produces an R^2 of 30.4% as against 25.4% for

the SR model, despite the SR model using the same market condition variables as per the GP approach whereas the HAR model only uses lagged volatility information as inputs. The GARCH model produces an R^2 of 29.71%, close to that of the HAR model.

4.2 Statistical Analysis

A series of statistical tests were undertaken to determine the significance of the results including three Diebold-Mariano tests introduced by [4] to test the equality of forecast accuracy between two models. The tests relate prediction error to some very general loss function and analyse the loss differential derived from errors produced by two competing models. The three tests include an asymptotic test that corrects for series correlation and two exact finite sample tests based on the sign test and the Wilcoxon signed-rank test. The last two sign-based tests in, particular, works well for small samples.

In this application, the differential loss is defined as the difference of the squared forecast errors from two competing models. The differential loss series (d_i) are calculated in Equation 11, where $Predicted_i^a$ is the predicted value by model a and $Predicted_i^b$ is the one from model b for the ith RV observation ($i = 1, \cdots, T$). Under the null hypothesis, the two competing models give equally accurate results. The alternative hypothesis is that two prediction models do not give equally accurate results. Three test statistics are shown in [4] to follow a standard normal distribution. The null hypothesis will be rejected at the 5 percent significant level if $|S| > 1.96$.

Table 4: Out-of-sample Forecast Error Measure

Model	MAE	MAPE	RMSE	R^2
HAR	0.014306	0.158370	0.017351	30.42%
GARCH	0.014411	0.160022	0.017442	29.71%
SR	0.014189	0.154038	0.017488	25.39%
GP-avg	0.013060	0.142854	0.016081	39.32%
Relative Change (GP vs HAR)	-8.71%	-9.80%	-7.32%	29.26%
Relative Change (GP vs GARCH)	-9.37%	-10.73%	-7.80%	32.35%
Relative Change (GP vs SR)	-7.96%	-7.26%	-8.05%	54.86%

$$d_i = (RV_i - Predicted_i^a)^2 - (RV_i - Predicted_i^b)^2 \tag{11}$$

– The Asymptotic Test: According to the Central Limit Theorem, when the sample size is large, the sample mean of the loss differential approximately follows a normal distribution with constant mean and variance. $\bar{d}$ is the sample mean and $\widehat{avar_{\bar{d}}}$ is the estimate of the asymptotic long-run variance of $\sqrt{T}\,\bar{d}$. In this application the forecast is only one step ahead, therefore no correlation adjustment is needed and $\widehat{avar_{\bar{d}}}$ is calculated as the variance of loss differential series. The statistic test is as per Equation 12.

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{\widehat{avar_{\bar{d}}}}{T}}} \sim_A N(0,1) \tag{12}$$

– The Sign Test: When the sample size is small, a finite sample test such as the sign test can be conducted. The sign test statistic is constructed in Equation 13, where $I(d_i)$ is one for $d_i > 0$ and otherwise zero.

$$S_2 = \frac{\sum_{i=1}^{T}[I(d_i)] - 0.5T}{\sqrt{0.25T}} \sim_A N(0,1) \tag{13}$$

– Wilcoxon's Signed Test: The test statistic is as per Equation 14, where $rank(|d_i|)$ is the rank of the absolute values of loss differential series.

$$S_3 = \frac{\sum_{i=1}^{T}[I(d_i) * rank(|d_i|)] - 0.5T}{\sqrt{0.25T}} \sim_A N(0,1) \tag{14}$$

Diebold-Mariano tests undertaken on a pairwise basis for the three competing models on the full out-of-sample time period are in Table 4. The resulting statistics are provided in Table 5. The null hypothesis, that two models give equal results in terms of forecasting accuracy, will be rejected at a 5 percent level if the relevant reported test statistic $|X| > 1.96$.

4.2.1 Summary of Statistical Results

The results from all three statistical tests give consistent results, that the prediction from GP is significantly different, from that produced by the HAR, GARCH and SR model, and as already seen in Table 4, the GP forecasts produced lowest error measures (and highest R^2) among the competing models. The null hypothesis of no difference in the predictive accuracy between the methods, is rejected by three Diebold-Mariano tests at a 5 percent level for GP model against the HAR, GARCH and SR model. All three statistical tests also consistently confirm that there is no statistical difference in the predictions from three benchmark models although one model is better than the other in the tested error measures.

4.3 Information Encompassing Tests

These tests are used to determine whether one of a pair of forecasts contains all the useful information for a forecast, or conversely, does one forecast contain additional information not captured by the other. In this case, use of a combination of the forecasts can produce a better forecast than either alone.

The forecast information encompassing tests are performed using regression analysis on the full out-of-sample time period and the results are displayed in Table 6.

$$RV = \alpha + \beta Predicted \tag{15}$$

Table 5: Diebold-Mariano Tests

		HAR	GARCH	SR	GP-avg
X1	HAR		-0.4731	-0.3303	3.3858
X2			-0.0880	0.7924	2.5533
X3			-0.5559	0.2809	3.4519
X1	GARCH	-0.4731		-0.1054	3.4055
X2		-0.0880		0.7924	3.6098
X3		-0.5559		0.8545	3.6776
X1	SR	-0.3303	-0.1054		4.1235
X2		0.7924	0.7924		2.7294
X3		0.2809	0.8545		3.3744
X1	GP-avg	3.3858	3.4055	4.1235	
X2		2.5533	3.6098	2.7294	
X3		3.4519	3.6776	3.3744	

This table gives D-M tests including the asymptotic test (X1), sign test (X2) and Wilcoxon's signed test (X3) on the full out-of-sample period. The null hypothesis that two models give equal accuracy results will be rejected at 5 percent significant level if $|X| > 1.96$.

Initially, a single factor analysis is performed for each model, where RV is the dependent variable and the prediction from each model is the explanatory variable as in Equation 15. The results from this are reported in Table 6. In evaluating these results it is important to distinguish between bias and predictive accuracy. In this single factor analysis, the prediction is unbiased only if $\alpha = 0$ and $\beta = 1$. The predictive power is indicated by R^2. A higher R^2 means higher predictive power. Ideally, we seek a forecast with low residual error and high R^2 [5]. While it might appear that bias is always undesirable, a biased forecast can still have predictive utility and conversely an unbiased forecast is of little use if the forecast errors produced by it are large.

The coefficients fitted in the single factor regression analysis in Table 6 shows that forecast results from the benchmark models (HAR, GARCH and SR) are closer to an unbiased prediction than those produced by GP. The intercept α are very close to zero and the coefficient for the model prediction, β are closer to one in the HAR model. In the case of GP, α is significant as it is not zero at the 5 percent level and the β value of 1.2813 is significantly higher than one. However, indicated by R^2 the predictive power from GP is much higher than that of the other models and hence it has significant utility despite its bias.

The second group of Information Encompassing Tests adds an extra prediction result from another model to the right-hand side of Equation 15 as a second regressor. An increased adjusted R^2 indicates that the first model can not subsume the second model and therefore that the second model provides extra predictive power. In other words, we are testing whether adding the prediction result from a second model as an extra explanatory factor can further improve the prediction result.

The adjusted R^2 from the regression of RV against the prediction from GP is found to be 38.84%, higher than the values for HAR, GARCH and SR. The adjusted R^2 values for the regression when both GP and one of the benchmark models, HAR / GARCH / SR forecasting results are used as regressors for RV increase to 40.20%, 39.24%, and 44.38% respectively. This indicates that the prediction from GP does not fully subsume the prediction from the benchmark models and suggests that a joint forecast from a hybrid of GP and the benchmark HAR, GARCH and SR models could potentially produce a higher predictive power.

From the empirical results, GP produces better forecasts than the benchmark models. There are two plausible reasons for this. First, GP takes account of market conditions (as inputs) in forecasting RV. Second, GP permits the use of non-linear functional forms between the RV and market conditions.

Table 6: Out-of-sample Forecast Information Encompassing Test

	HAR	GARCH	SR	GP-avg
α	-0.0048	-0.0132	0.0028	-0.0302
p-Value	0.7302	0.3863	0.8471	0.0346
β	1.0043	1.0876	0.9545	1.2813
p-Value	0.0000	0.0000	0.0000	0.0000
R^2	30.42%	29.71%	25.39%	39.32%
Adj-R^2	29.87%	29.16%	24.80%	38.84%

Adj-R^2	HAR	GARCH	SR	GP-avg
HAR		29.35%	29.33%	40.20%
GARCH	29.35%		28.69%	39.24%
SR	29.33%	28.69%		44.38%
GP-avg	40.20%	39.24%	44.38%	

The above two tables provides forecast Information Encompassing Test results on the out-of-sample time period. The first table gives a regression fitting result when regressing the RV on the predicted value from each of five models as shown in Equation 15. The second table gives the adjusted R^2 when extra predictive results from another model is added in to the right-hand side of the Equation 15 as a second regressor. An increased adjusted R^2 indicates that the first model can not subsume the second model and the second model provides extra predictive power.

It is found from analysis of the form of GP generated solutions that some variables including lagged values of RV, and average trading duration occurred frequently. The relationship between these factors and RV seem robust over time.

5 Conclusions

Forecasting daily returns volatility is crucial in finance. Traditionally, volatility is modeled using a time-series of lagged information only, an approach which is in essence atheoretical. Although the relationship of market conditions and volatility has been studied for decades, we still lack a clear theoretical framework to allow us to forecast volatility, despite having many plausible explanatory variables. This setting of a theory-weak environment suggests a useful role for powerful model induction methodologies such as Genetic Programming. This study forecasts one-day ahead realised volatility (RV) using a GP methodology that incorporates information on market conditions including trading volume, number of transactions, bid-ask spread, average trade duration and implied volatility. The forecasting result from GP is significantly better than that produced by the heterogeneous autoregressive model (HAR), the generalized autoregres-

sive conditional heteroscedasticity (GARCH) and a linear stepwise regression (SR) model. Further, the regression-based Information Encompassing Tests show that the forecasts from benchmark models contain information not captured by GP, which indicates that a combination forecast from GP and conventional models could potentially improve forecast performance further. This is left for future work.
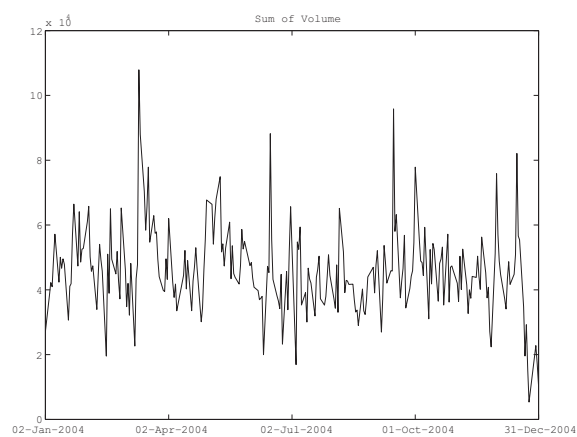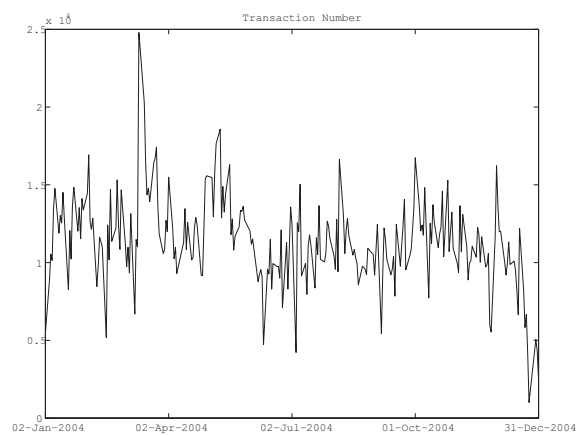
Acknowledgment
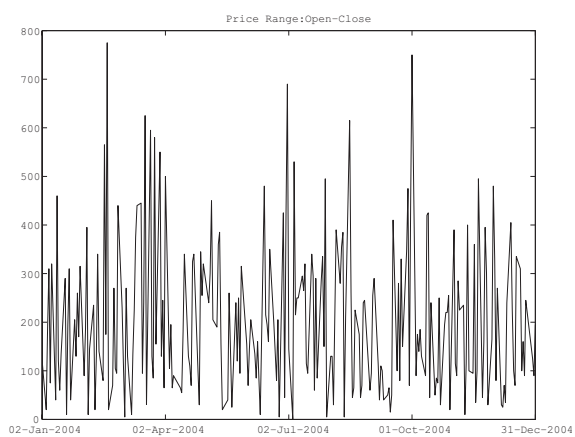
Figure 5: Absolute Difference of Open to Close

Figure 6: Absolute Difference of High to Low

Figure 7: Daily Squared Return

Figure 8: Total of Trading Volume

Figure 9: Average 5-minute Trading Volume

Figure 10: Transaction Number
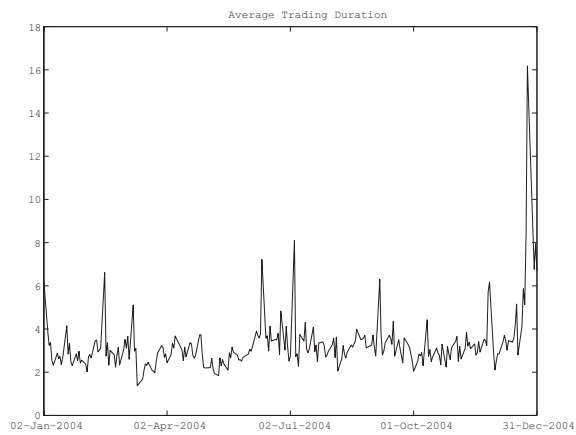
Figure 11: Average Trading Duration
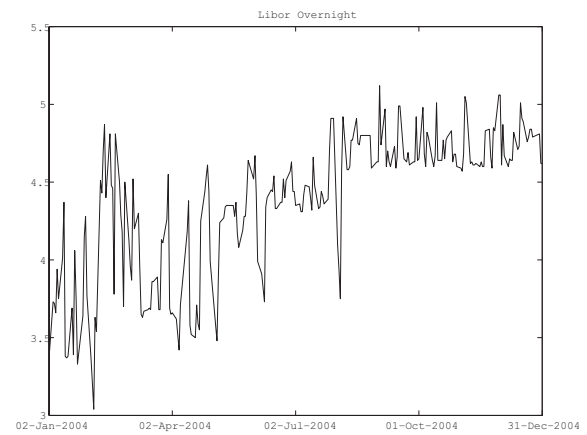
Figure 14: Daily Overnight LIBOR
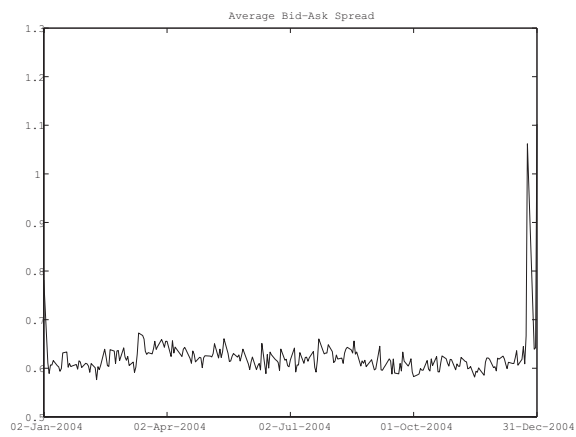
Figure 12: Average Bid-Ask Spread
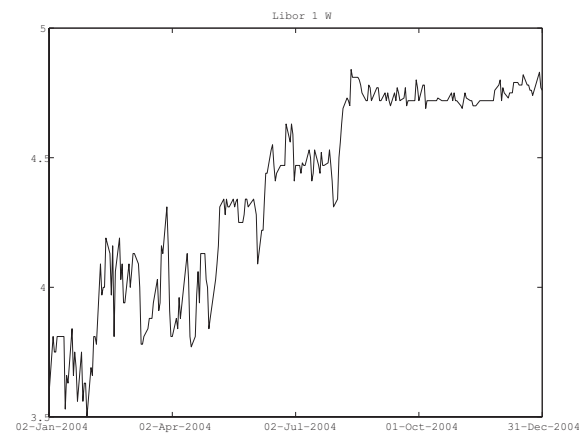
Figure 15: Daily 1-week LIBOR
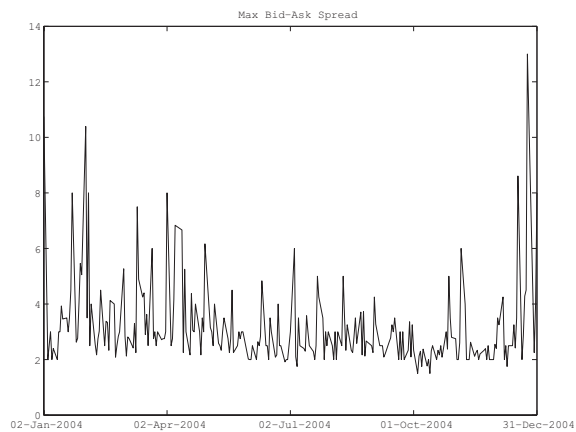
Figure 13: Max Bid-Ask Spread
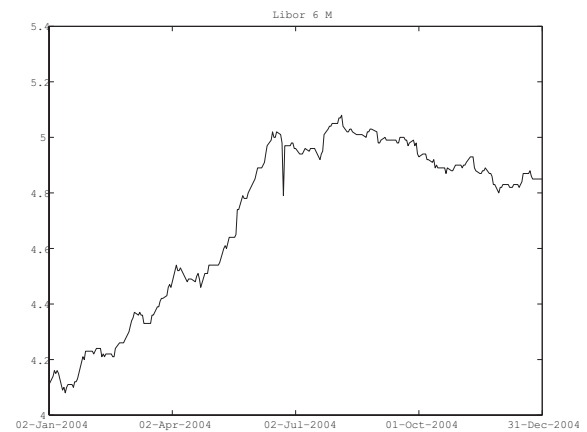
Figure 16: Daily Six-month LIBOR

References

[1] R.F. Engle. GARCH 101: The use of ARCH/GARCH models in applied econometrics. Journal of Economic Perspectives, vol. 14, pp. 157-168, 2001.

[2] N. Shephard. Stochastic Volatility. Economics Group, Nuffield College, University of Oxford, Economics Papers, 2005.

[3] W.K.H. Fung and D.A. Hsieh. Empirical Analysis of Implied Volatility: Stock, Bonds and Currencies, presented at the 4th Annual Conference of the Financial Operations Research Center, University of Warwick, Coventry, England, 1991.

[4] F.X. Diebold and R. Mariano. Comparing Predictive Accuracy. Journal of Business and Economic Statistics, vol. 13, pp. 253-265, 1995.

[5] S.H. Poon and C.W.J. Granger. Forecasting Volatility in Financial Markets: A Review. Journal of Economic Literature, vol. XLI, pp. 478-539, 2003.

[6] J.R. Koza. Genetic Programming. Massachusetts: MIT Press, 1992.

[7] J. R. Koza. Genetic Programming II. Massachusetts: MIT Press, 1994.

[8] Poli R., Landon W.B. and McPhee N.F (2008) A field guide to genetic programming, Published via http://lulu.com and freely available: http://www.gp-field-guide.org.uk

[9] S.-H. Chen and C.H. Yeh. Using Genetic Programming to Model Volatility in Financial Time Series: the Case of Nikkei 225 and S&P 500. in Proc. of the 4th JAFEE International Conference on Investments and Derivatives(JIC'97), Aoyoma Gakuin University, Tokyo, Japan, July 29-31, 1997, pp.288-306.

[10] R.F. Engle. The Econometrics of Ultra High Frequency Data. Econometrica, vol. 68, pp. 1-22, 2000.

[11] G. Zumbach, O.V. Pictet and O. Masutti. (2001) Genetic Programming with Syntactic Restrictions Applied to Financial Volatility Forecasting. Olsen & Associates Working Paper, [online] Paper No. GOZ.2000-07-28. Available: http://ssrn.com/abstract=269189 [15 Dec. 2006].

[12] C.J. Neely and P.A. Weller Predicting Exchange Rate Volatility: Genetic Programming Versus GARCH and RiskMetrics^{TM}. The Federal Reserve Bank of St. Louis, May/June, 2002.

[13] I. Ma, T. Wong , T. Sankar and R. Siu. Forecasting the Volatility of a Financial Index by Wavelet Transform and Evolutionary Algorithm. in Proc. of the 2004 IEEE International Conference on Systems, Man and Cybernetics, pp. 5824-5829, 2004.

[14] S. B. Hamida and R. Cont. Recovering Volatility from Option Prices by Evolutionary Optimization. Journal of Computational Finance, vol. 8, No. 4, Summer 2005.

[15] W. Abdelmalek, S. B. Hamida and F. Abid. (2009). Selecting the Best Forecasting-Implied Volatility Model Using Genetic Programming. Journal of Applied Mathematics and Decision Sciences, [online]. vol. 2009, Article ID 179230. Available: http://www.hindawi.com/journals/ads/2009/179230/ [17 Sep.2011].

[16] H.M. Wu and W.C. Guo. Asset Price Volatility and Trading Volume with Rational Beliefs. Economic Theory, vol. 23, no. 4, pp. 795-829, Jun. 2004.

[17] H.C. Chena and J. Wu. Return Volatility and the Intraday Behavior of Market Liquidity without Market Makers: Evidence from the Taiwan Futures Market. International Research Journal of Finance and Economics, Issue 17, pp. 117-128, Jul. 2008.

[18] S.J. Taylor. Forecasting the Volatility of Currency Exchange Rates. International Journal of Forecasting, vol. 3, pp. 159-170, 1987.

[19] L.R. Glosten, R. Jagannathan and D.E. Runkle. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance, vol. 48, no. 5, pp. 1779-1801, 1993.

[20] M. Ammann, D. Skovmand and M. Verhofen. Implied and Realized Volatility in the Cross-Section of Equity Options. International Journal of Theoretical and Applied Finance, vol. 12, issue 6, pp. 745-765, 2009.

[21] G.H.K. Wang and J. Yau. Trading Volume, Bid-Ask Spread, and Price Volatility in Futures Markets. Journal of Futures Markets, vol. 20, issue 10, pp. 943-970, 2000.

[22] M. McAleer and M.C. Medeiros. Realized Volatility: A Review. Econometric Reviews, vol. 27(1-3), pp. 10-45, 2008.

[23] T.G. Andersen, T. Bollerslev, F.X. Diebold and P. Labys. Modeling and Forecasting Realized Volatility. Econometrica, vol. 71, pp. 529-626, 2003.

[24] T.G. Andersen and T. Bollerslev. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. International Economic Review, vol. 39, no. 4, pp. 885-905, 1998.

[25] N.M.P.C. Areal and S.J. Taylor. The Realized Volatility of FTSE 100 Futures Prices. Journal of Futures Markets, vol. 22, issue 7, pp. 627-648, 2000.

[26] D.D. Thomakos and T. Wang. Realized Volatility in the Futures Markets. Journal of Empirical Finance, vol. 10, pp. 321-353, 2003.

[27] F. Corsi. A Simple Long Memory Model of Realized Volatility. Working Paper, University of Southern Switzerland, vol. 71, pp.529-626, 2003.

[28] Y. Aït-Sahalia and L. Mancini. Out of Sample Forecasts of Quadratic Variation. Journal of Econometrics, vol. 147, pp. 17-33, 2008.

[29] L. Xiao. Realized Volatility Forecasting: Empirical Evidence from Stock Market Indices and Exchange Rates. Applied Financial Economics, vol. 23, pp. 57-69, 2013.

[30] F. Corsi, S. Miittnik, Ch. Pigorsch and U. Pigorsch. The Volatility of Realized Volatility. Econometric Reviews, vol. 27, pp. 46-78, 2008.

[31] T.G. Andersen, T. Bollerslev and F.X. Diebold. Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. Review of Econometrics and Statistics, vol. 89, pp. 701-720, 2007.

[32] F. Corsi. A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics, vol. 7, no. 2, pp. 174-196, 2009.

[33] A.M. Fuertes, M. Izzeldin and E. Kalotychou. On Forecasting Daily Stock Volatility: The Role of Intraday Information and Market Conditions. International Journal of Forecasting, vol. 25, pp. 259-281, 2009.

[34] N.K. Léon. (2008) The Effects of Interest Rates Volatility on Stock Returns and Volatility: Evidence from Korea. International Research Journal of Finance and Economics, [online]. ISSN 1450-2887, issue 14. Available: http://www.eurojournals.com/finance.htm[1 Jan. 2012].

[35] N. Zafar, S.F. Urooj and T.K. Durrani. (2008) Interest Rate Volatility and Stock Return and Volatility. European Journal of Economics, Finance and Administrative Sciences, [online]. ISSN 1450-2887, Issue 14. Available: http://www.eurojournals.com [1 Jan. 2012].

[36] S. Rahman, C.F. Lee and K.P. Ang. Intraday Return Volatility Process: Evidence from NASDAQ Stocks. Review of Quantitative Finance and Accounting, vol. 19, pp. 155-180, 2002.

[37] E. Kalotychou and S.K. Staikouras. Volatility and Trading Activity in Short Sterling Futures. Applied Economics, vol.38, no. 9, pp. 997-1005, 2006.

[38] C.M. Jones and G. Kaul. Transactions, Volume, and Volatility. The Review of Financial Studies, vol. 7, no. 4, pp. 631-651, 1994.

[39] L.R. Glosten, R. Jagannathan and D.E. Runkle. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance, vol. 48, no. 5, pp. 1779-1801, 1993.

[40] A.F. Darrat, S. Rahman and M. Zhong. Intraday Trading Volume and Return Volatility of the DJIA Stocks: A Note. Journal of Banking & Finance, vol. 27, pp. 2035-2043, 2003.

[41] C. Brooks. Predicting Stock Index Volatility: Can Market Volume Help? Journal of Forecasting, vol. 17, pp. 59-80, 1998.

[42] Research & Planning Department.(2006, Oct.) Global Exchange Derivative Volumes - A Statistical Overview. Exchange Newsletter (Hong Kong, China). [Online] Available: http:www.hkex.com.hk/eng/newsconsul/newsltr/2011newsletter.htm [15 Jul. 2013].

[43] R.F. Engle and G.M. Gallo. A Multiple Indicators Model for Volatility Using Intra-daily Data. Journal of Econometrics, vol. 131, pp. 3-27, 2006.

[44] G. Donaldson and M. Kamstra. Volatility Forecasts, Trading Volume, and the ARCH versus Option-Implied Volatility Trade-off. Journal of Financial Research, Southern Finance Association & Southwestern Finance Association, vol.28, no. 4, pp.519-538, 2005.

[45] C. Christiansen, M. Schmeling and A. Schrimpf. A Comprehensive Look at Financial Volatility Prediction by Economic Variables. BIS Working Paper,no. 374, pp. 41-45, 2008.

[46] S. Li and Q. Yang. The Relationship between Implied and Realized Volatility: Evidence from the Australian Stock Index Option Market. Rev. Quant Finance Account, vol. 32, pp. 405-419, 2009.

[47] J Shu and J.E. Zhang. The Relationship between Implied and Realized Volatility of S&P 500 Index. WILMOTT Magazine, Technical Article 4, pp. 83-91, 2003.

[48] B. Blair, S.H. Poon and S. J. Taylor. Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High Frequency Index Returns. Journal of Econometrics, vol. 105 (1), pp. 5-26, 2001.

[49] B.J. Christensen and N.R. Prabhala. The Relation Between Implied and Realized Volatility. Journal of Financial Economics, vol. 50, pp. 125-150, 1998.

[50] L. Canina and S. Figlewski. The Informational Content of Implied Volatility. The Review of Financial Studies, vol. 6, no. 3, pp. 659-681, 1993.

Dr. Yin completed the undergraduate study of Economics in Shaanxi University of Finance & Economics, P.R. China. She received MSc of Quantitative Finance and Ph.D degree from the Michael Smurfit Graduate Business School, University College Dublin, Ireland. She is currently a researcher in the Graduate Business School, University College Dublin and a senior risk analyst in Allied Irish Bank. Her main research interests include applying evolutionary computation in financial modelling, financial market microstructure, financial derivatives hedging, credit risk modelling. She has presented papers at international conferences including GECCO (2007 Genetic and Evolutionary Computation Conference) and CEC (2015 IEEE Congress on Evolutionary Computation).

Dr. O'Sullivan is a lecturer of finance and director of the MSc in Quantitative Finance at the Michael Smurfit Graduate Business School, University College Dublin. His primary research interests are numerical methods for derivatives pricing, derivatives markets, interest rate models and the application of natural computing algorithms in finance. He has published works in journals such as Quantitative Finance, the International Journal of Theoretical and Applied Finance and the Journal of Computational Finance. He presented papers at international conferences including the European Finance Association, Quantitative Methods for Finance and Computational Financial Econometrics.

Prof. Brabazon's research interests concern the development of natural computing algorithms and their application to real-world problems. He is a co-founder and co-director of the Natural Computing Research & Applications Group at UCD (see http://ncra.ucd.ie). Has published in excess of 200 peer-reviewed journal and conference papers. He has also authored or edited fourteen books including 'Biologically Inspired Algorithms for Financial Modelling', 'Foundations in Grammatical Evolution for Dynamic Environments' and 'Natural Computing in Computational Finance' (Volumes I, II, III and IV), and 'Natural Computing Algorithms'. Prof. Brabazon completed his undergraduate studies at UCD and subsequently undertook graduate degree studies at a number of universities including Trinity College Dublin, Heriot-Watt, Kingston and Stanford.