UDC 336.77, 519.615
DOI: 10.31891/2307-5740-2021-292-2-4

MAKSYM BONDARENKO
ORCID ID: 0000-0002-4792-6420
e-mail: maksym_bondarenko@outlook.com
VICTOR BONDARENKO
ORCID ID: 0000-0003-1663-4799
e-mail: vicnbondarenko@gmail.com
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»

MODELING RELATION BETWEEN ATM LOCAL AND IMPLIED VOLATILITY FOR MICROSOFT STOCKS

In this work simple linear and polynomial regression to model the relation between at-the-money (ATM) implied and at-the-money local volatility of Microsoft stocks has been applied.

Local volatility is extracted from the set of Vanilla option prices on Microsoft stocks by assuming that Microsoft stock price follows Dupire local volatility process. ATM Local volatility is then used in linear regression predictor while implied volatility is a resulting variable.

The model is validated by predicting out-of-sample implied volatility with local volatility. The statistical significance and predictive ability of such model have been measured and autocorrelation tendencies have been studied.

The conclusion that assumptions to use linear regression are held has been made. No autocorrelation tendencies were discovered in the time series.

Finally, the conclusion that both the 1st and the 3rd order linear regression models demonstrate good predictive ability of local volatility over out-of-sample implied volatility has been made. None of the models proves statistical significance of local volatility as a predictor of the implied volatility but both can be actually used for practical purpose as they predict well out-of-the-sample implied volatilities. This is an important practical result as it means that complex non-linear relationship between implied and local volatilities formalized by Dupire can actually be reduced to simplier linear relationship that demonstrates reasonable discrepancies.

Despite the 3rd order model fits the data better, but for the reasons of overfitting in general it's safer to apply the 1st order model as it demonstrates more stable predictions over datasets with jumps.

Keywords: regression; implied volatility; calibration; Microsoft stocks; Dupire model; statistical significance; predictive ability.

БОНДАРЕНКО М. В.,
БОНДАРЕНКО В. М.
Національний технічний університет України
«Київський Політехнічний Інститут імені Ігоря Сікорського»

МОДЕЛЮВАННЯ СПІВВІДНОШЕННЯ МІЖ ЛОКАЛЬНОЮ ТА НЕЯВНОЮ ВОЛАТИЛЬНІСТЮ «У ГРОШАХ» ДЛЯ АКЦІЙ MICROSOFT

Проста лінійна та поліноміальна регресії застосовані для моделювання співвідношення між неявною волатильністю «у грошах» (ATM) та локальною волатильністю акцій Microsoft. Локальна волатильність визначається із набору цін на опціони Vanilla по акціях Microsoft у припущенні, що ціна акцій Microsoft відповідає процесу локальної волатильності Дюпіра. Далі локальна волатильність «у грошах» (ATM) використовується в прогнозі лінійної регресії, тоді як неявна волатильність є результуючою змінною. Перевірку моделі виконано шляхом прогнозу неявної волатильності поза вибіркою через локальну волатильність. Також, вимірюються статистична значимість і прогнозуюча здатність такої моделі та вивчаються тенденції автокореляції. Зроблено висновок, що припущення щодо використання лінійної регресії виконані. Жодних тенденцій автокореляції у часових рядах не виявлено. Нарешті, зроблено висновок, що обидві моделі лінійної регресії 1-го та 3-го порядку демонструють хорошу прогностичну здатність локальної волатильності для неявної волатильності поза вибіркою. Жодна з моделей не доводить статистичної значущості локальної волатильності, як предиктора неявної волатильності, але обидві вони насправді можуть бути використані для практичних цілей, оскільки добре передбачають неявну волатильність поза вибіркою. Маємо важливий практичний результат, оскільки складний нелінійний взаємозв'язок між неявною та локальною волатильностями, формалізований Дюпіром, насправді може бути зведений до спрощеного лінійного зв'язку, який демонструє розумні розбіжності. Незважаючи на те, що модель третього порядку краще підходить для даних, з причин перевизначення, загалом безпечніше застосовувати модель 1-го порядку, оскільки вона демонструє більш стабільні прогнози щодо наборів даних зі стрибками.

Ключові слова: регресія, неявна волатильність, калібровка, акції Microsoft, модель Дюпіра, статистична значимість, предикативна здатність.

Introduction. The relevance of the study is due to the need for high-quality forecasting of market volatility by different market participants and the possibility of using a modern model of local volatility [1] to achieve this goal. A number of researches, including Christensen [2], consider the "at-the-money" (ATM) implied volatility to be a good measure of the expected variation in the stock price. That's why forecasting such a measure can have practical application in arbitrage and evaluation of options. Christensen's main achievement is the calibration of implicit volatility based on non-intersecting data over a long period of time.

In this paper, a shorter period of time is applied, which does not allow the use of only non-intersecting data. Calibrating the matrix of local volatility $\sigma(T,K)$ at different points in time in the past, time series of local volatility has been constructed. For input data, the prices of Vanilla options (Call) on Microsoft stocks from the NASDAQ exchange during the crisis period of the pandemic from January to April 2020 were selected [3].

The purpose of the study is to assess the predictive power of linear models based on time series of at-the-money local volatility [1] derived from Microsoft stocks price data.

Theoretical basics. Local volatility and implied volatility [4] matrices are calibrated from prices of Vanilla options (Call and Put) available on the market. Such matrices have to be calibrated at different calibration dates in the past. The underlying stock price is supposed to follow Dupire model dynamics [1]. Our goal is to model the relation between at-the-money (ATM) implied and at-the-money local volatility of Microsoft stocks. The 1st and the 3rd order polynomial linear regressions are used to see if at-the-money local volatility can predict the implied volatility and the results on out-of-the-sample (OOS) validation points within two different datasets are verified.

A set of prices of Vanilla options on MSFT stocks from NASDAQ has been used. Option prices are available for multiple trading dates between January and April 2020 [3]. As input data is sparse and maturities/strikes differ between calibration dates, linear interpolation on maturity axis is used to obtain the same grid for each of calibration dates and fill missing values. The grid is represented by these fractions of the year:

$$Maturities\ grid\ (N=6):\ [\ 0.05, 0.1, 0.15, 0.2, 0.25, 0.4\ ] \quad (1)$$

Given N – number of maturities and M – number of strikes, $N\ X\ M$ matrix is firstly calibrated from the set of options for each of K calibration dates using genetic algorithm of optimization by Cerf [5] to handle the ill-posed character of Dupire calibration problem as shown in Bondarenko [6]. Hence, K local volatility matrices of $N\ X\ M$ dimension have been obtained.

Because of its practical significance but also to reduce the dimension of the problem, only ATM local volatilities are extracted. Thus, $N\ X\ K$ matrix of local volatilities without strike axis is obtained. Our term structure corresponds to calibration dates from Table 1 and Table 2 below.

Then K implied volatility matrices are calibrated using Newton–Raphson algorithm of optimization [7].

Christensen suggests that implied volatility is better predictor of realized volatility than past realized volatility [2]. The task of this article is to model implied volatility in relation with ATM local volatility before making any predictions of realized one.

The relation between implied and local volatility has been described by Derman, Kani and Zou [8] and Gatheral [9] as equation (2). In the equation $\sigma(T,K)$ is local volatility function at maturity T and strike K and Σ is implied volatility function at same maturity and strike, and d1 and d2 – known Black-Scholes parameters:

$$\sigma(T,K) = \sqrt{\frac{\Sigma^2+2\Sigma T(\frac{d\Sigma}{dT}+rK\frac{d\Sigma}{dK})}{1+2d_1K\sqrt{T}\frac{d\Sigma}{dK}+K^2T\left(d_1d_2\left(\frac{d\Sigma}{dK}\right)^2\right)+\Sigma\frac{d^2\Sigma}{dK^2}}} \quad (2)$$

However, this relationship is nonlinear and has some conditions as differentiability of implied volatility surface at these strikes and maturity. So, the presented approach is to model the implied volatility with local volatility linearly using the regression model and see if this linear simplification can be statistically significant.

Linear (3) and polynomial (4) regressions are chosen as instruments to model the implied volatility (dependent variable) w.r.t. Local Volatility at K corresponding calibration dates at each of N maturities. It has been proven earlier that calibration date is neither a good predictor of the local volatility nor there are any autocorrelation tendencies observable for current time-series, so the calibration date t is not used as second factor to our regression model and the fact that Local Volatility is a function of t is ignored.

$$\widehat{ImpliedVol}_T(t) = a_T + b_T * LocalVolATM_T(t) \quad (3)$$

$$\widehat{ImpliedVol}_T(t) = a_T + b_T * LocalVolATM_T(t) + c_T * LocalVolATM_T(t)^2 + d_T * LocalVolATM_T(t)^3 \quad (4)$$

Regression assumptions are verified using graphical analysis of distribution plots of residuals (e.g., in Figure 1 below).

Further, r-squared, t and F-statistics, and out-of-sample MSE (mean squared error) are used to evaluate the quality of regression model. Durbin-Watson test is conducted to determine the presence of autocorrelation (may be caused by overlapping) in our data.

Simple linear and polynomial models are validated on two different sets of calibration/validation dates. MSE measure (5) and analysis of plots is used to measure the quality of predictions and validate obtained results. Cross validation shows if the model can successfully predict the implied volatility at intermediary OOS dates (Table 1) while predictive validation uses the implied volatility at extrapolated OOS calibration dates (Table 2).

$$MSE_T = \sum_t \frac{abs(ImpliedVol_T(t) - \widehat{ImpliedVol}_T(t))}{ImpliedVol_T(t)} \quad (5)$$

Table 1

Cross validation dataset

Calibration Dates	01/03/2020	01/10/2020		01/24/2020	01/31/2020	02/07/2020		
Validation dates (OOS)			01/17/2020				02/14/2020	02/21/2020
Spot MSFT	158.6200	161.3400	167.1000	165.0400	170.2300	183.8900	185.3500	178.5900
Calibration Dates	02/28/2020	03/06/2020		03/20/2020	03/27/2020	04/03/2020		
Validation dates (OOS)			03/13/2020				04/09/2020	04/17/2020
Spot MSFT	162.0100	161.5700	158.8300	137.3500	149.7000	153.8300	165.1400	178.6000

Two runs on each set (basically recalibrating local volatility twice) have been done in order to reflect on ill-posedness of Dupire optimization problem [10]: no unique solution exists. So, two runs allow us to use two different inputs (local volatility matrices) to regression model and potentially obtain different results.

Table 2

Predictive validation dataset

Calibration Dates	01/03/2020	01/10/2020	01/17/2020	01/24/2020	01/31/2020	02/07/2020	02/14/2020	02/21/2020
Validation dates (OOS)								
Spot MSFT	158.6200	161.3400	167.1000	165.0400	170.2300	183.8900	185.3500	178.5900
Calibration Dates	02/28/2020	03/06/2020						
Validation dates (OOS)			03/13/2020	03/20/2020	03/27/2020	04/03/2020	04/09/2020	04/17/2020
Spot MSFT	162.0100	161.5700	158.8300	137.3500	149.7000	153.8300	165.1400	178.6000

Numerical results

ANOVA (analysis of variance) is conducted to verify the assumptions necessary to use the linear regression (Figure 1).

Cross-validation set **Predictive validation set**

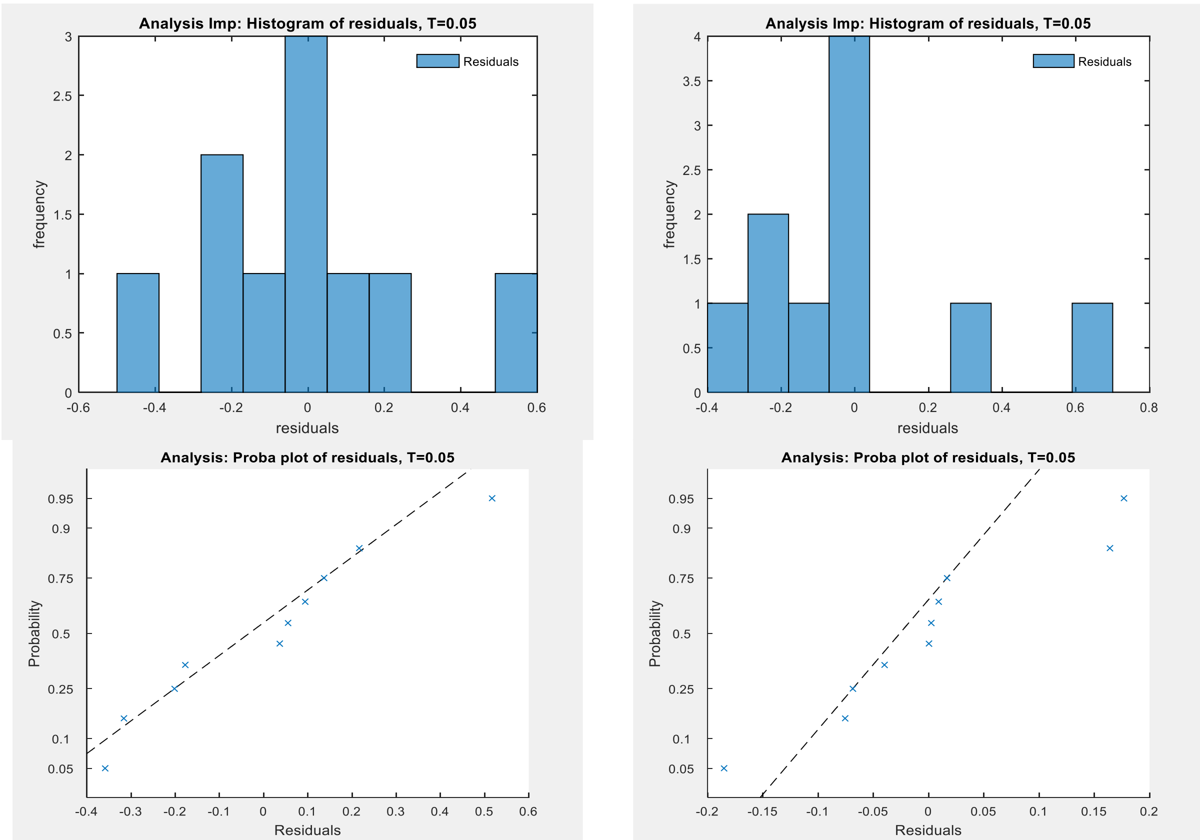

Fig. 1. ANOVA for cross-validation and predictive sets at T=0.05

Let's take a look at numerical results obtained for the first dataset.

For cross validation set Figure 2 and Figure 3 illustrate the dynamics of calibrated Implied volatility for the maturities T = 0.05 and T = 0.4 respectively; as well as the regression plots (modelled implied volatility w.r.t. local volatility) that have been built (the 1st and the 3rd order). For the second (predictive validation) set the same data are illustrated in Figure 4 and Figure 5. Subplots illustrate the dynamics of underlying MSFT stock price whose peaks correspond to volatility peaks as can be concluded.

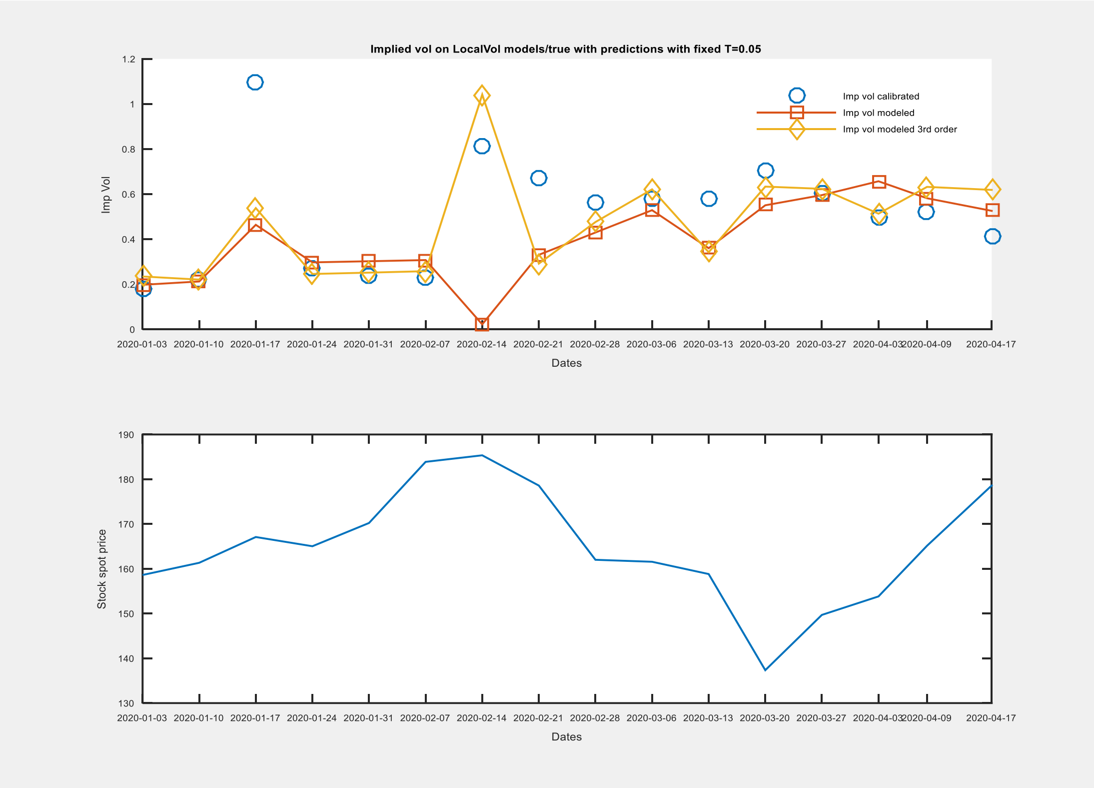

Fig. 2. Calibrated and modelled implied volatility values (w.r.t. local volatility) at T=0,05, Cross-Validation set

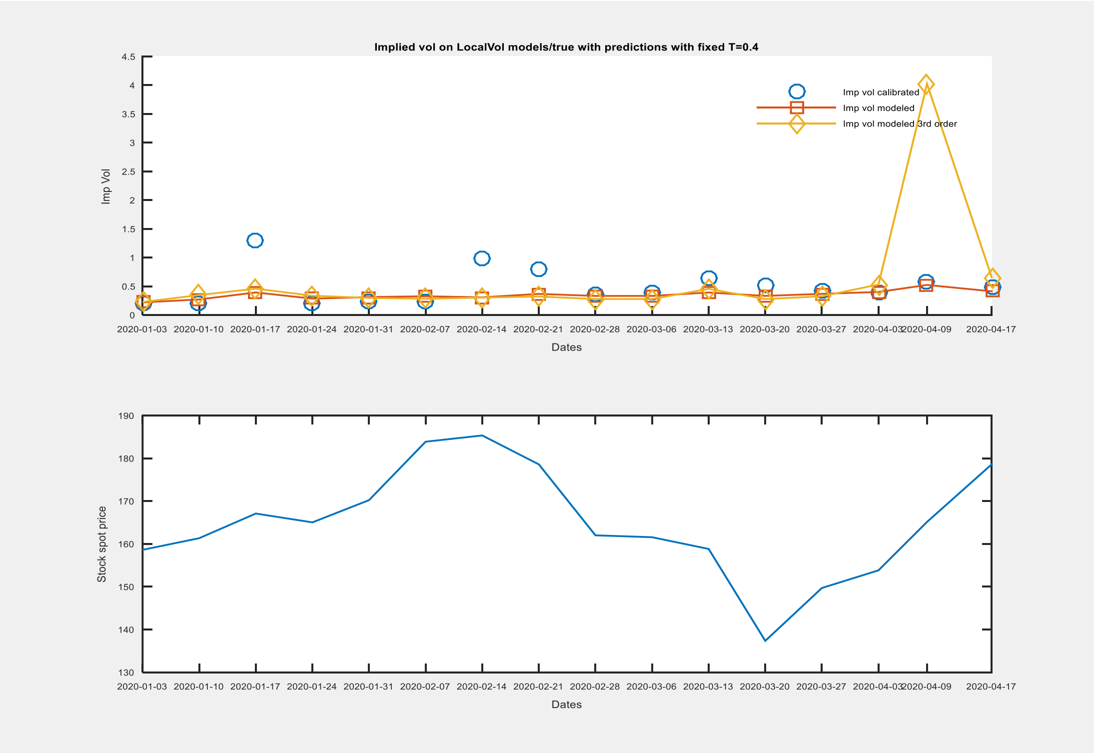

Fig. 3. Calibrated and modelled implied volatility values (w.r.t. local volatility) at T=0,4, Cross-Validation set

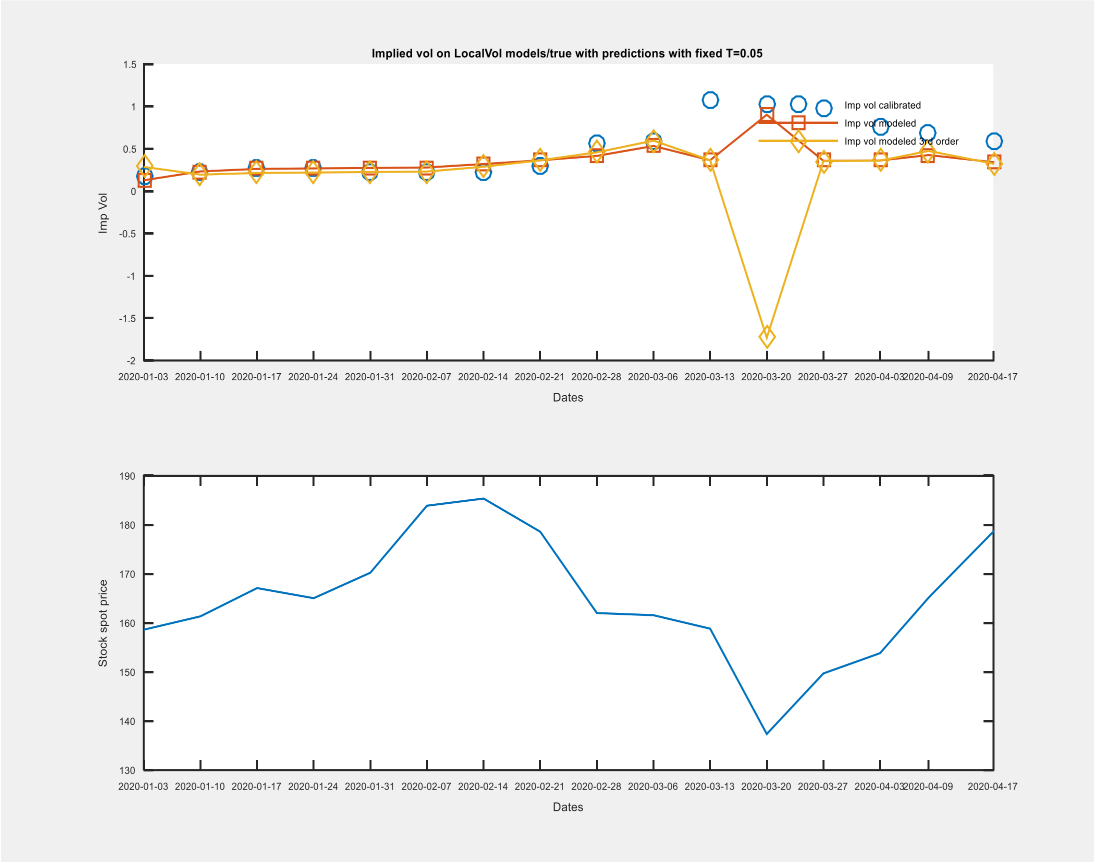

Fig. 4. Calibrated and modelled local volatility values (w.r.t. date) at T=0,05, Predictive Set

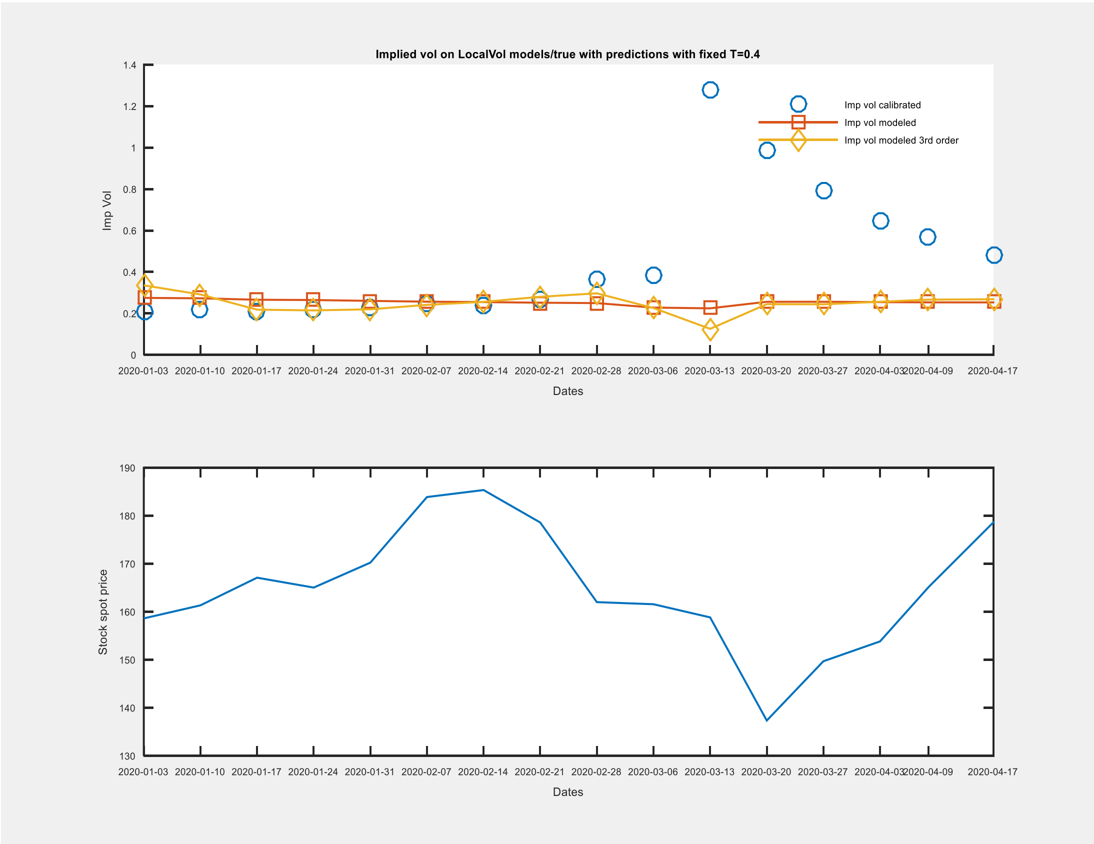

Fig. 5. Calibrated and modelled local volatility values (w.r.t. date) at T=0,4, Predictive Set

Figure 6 illustrates similarity between surfaces of local and implied volatility calibrated at $t = 02/21/2020$.

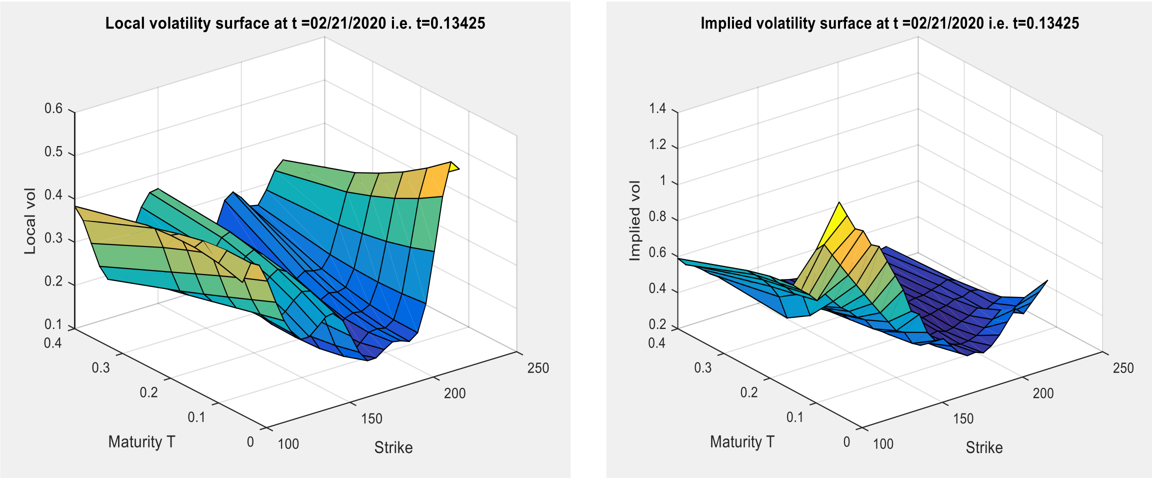

Fig. 6. Calibrated local and implied volatility surfaces at $t = 02/21/2020$

Table 3 illustrates the statistics of the 1st order regression model for two different calibration runs (different ATM local volatility as input) for both cross-validation and predictive sets.

Table 3

Stat the 1st order linear regression, Cross-Validation and Predictive sets

CROSS_VALIDATION SET								
RUN 1								
T	**Rsq**	**tStat**	**pValue**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.1407	-1.1446	0.28546	1.31	0.285	1.7293	0.2348	0.1401
0,1	0.2387	-1.5838	0.15191	2.51	0.152	1.2527	0.0374	0.1578
0,15	0.1274	-1.0809	0.31123	1.17	0.311	1.9153	0.4063	0.1426
0,2	0.2113	-1.4638	0.18139	2.14	0.181	1.3048	0.0421	0.0993
0,25	0.1756	-1.3056	0.22799	1.7	0.228	1.3692	0.0544	0.0807
0,4	0.1238	-1.0633	0.31867	1.13	0.319	1.2464	0.0240	0.1080
RUN 2								
T	**Rsq**	**tStat**	**pValue**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.0028	-0.14948	0.88487	0.0223	0.885	1.6225	0.1475	0.1604
0,1	0.1564	1.218	0.25791	1.48	0.258	2.2608	0.7892	0.2777
0,15	0.1394	1.1384	0.28789	1.3	0.288	2.3170	0.8807	0.2378
0,2	0.2799	1.7632	0.11587	3.11	0.116	0.116	0.8712	0.2477
0,25	0.3477	2.0652	0.072775	4.26	0.0728	3.0196	0.2161	0.2193
0,4	0.3803	2.2156	0.057573	4.91	0.0576	1.4051	0.0582	0.2167
PREDICTIVE SET								
RUN 1								
T	**Rsq**	**tStat**	**pValue**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.1294	-1.0903	0.30733	1.19	0.307	2.9044	0.3287	0.2406
0,1	0.0006	0.067196	0.94807	0.00452	0.948	1.3169	1.3169	0.1198
0,15	0.0432	0.60084	0.56458	0.56458	0.56458	1.9367	0.4014	0.1331
0,2	0.0029	0.15272	0.15272	0.0233	0.0233	2.2508	0.7988	0.1157
0,25	0.0016	0.11219	0.913	0.0126	0.0126	1.8571	0.3304	0.0986
0,4	0.0001	-0.024664	0.98093	0.000608	0.000608	2.5982	0.7200	0.1306
RUN 2								
T	**Rsq**	**tStat**	**pValue**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.1292	-1.0894	0.3077	1.19	0.308	2.4755	0.8897	0.2041
0,1	0.1376	-1.1298	0.29129	1.28	0.291	2.6690	0.6320	0.1288
0,15	0.0260	-0.46245	0.65608	0.214	0.656	2.0589	0.5116	0.1524
0,2	0.0986	-0.93565	0.37683	0.875	0.377	1.3139	0.0343	0.1555
0,25	0.1116	-1.0025	0.34547	1	0.34547	1.5255	0.0935	0.1455
0,4	0.0499	-0.64822	0.535	0.42	0.535	2.0082	0.4403	0.1686

Where:
T – is maturity expressed in fraction of the year;
Rsq – r-squared (coefficient of determination);
tStat – t-Statistics;

Fstat – F-statistics;

pValue – p-value;

DW – Durbin-Watson coefficient;

MSE – mean-squared error.

Table 4 illustrates the statistics of the 3rd order regression model for two different calibration runs (different ATM local volatility as input) for both cross-validation and predictive sets.

Table 4

Stat the 3rd order linear regression, Cross-Validation and Predictive sets

CROSS_VALIDATION SET						
RUN 1						
T	**Rsq**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.3923	1.29	0.36	1.7560	0.0588	0.2265
0,1	0.7065	4.81	0.0488	1.8745	0.2125	0.1989
0,15	0.6594	3.87	0.0745	3.4187	0.0944	0.1680
0,2	0.6000	3	0.117	1.9066	0.1590	0.1149
0,25	0.4520	1.65	0.275	1.9525	0.1500	0.0870
0,4	0.5126	2.1	0.201	1.5000	2.3162e-05	0.1092
RUN 2						
T	**Rsq**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.3481	1.07	0.43	2.8680	0.8911	0.1908
0,1	0.1586	0.377	0.773	2.3067	0.3559	0.2723
0,15	0.1481	0.348	0.793	2.3720	0.5608	0.2161
0,2	0.4145	1.42	0.327	2.3844	0.6267	0.2272
0,25	0.5770	2.73	0.137	3.3900	0.1063	0.2140
0,4	0.6696	4.05	0.0684	2.4047	0.5237	0.2023
PREDICTIVE SET						
RUN 1						
T	**Rsq**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.1996	0.499	0.697	3.0754	0.4848	0.5688
0,1	0.7592	6.31	0.0276	2.7530	0.8510	3.4680
0,15	0.2948	0.836	0.521	2.5725	0.6917	1.7081
0,2	0.2673	0.729	0.571	2.6864	0.9824	1.9369
0,25	0.2488	0.662	0.605	2.2339	0.4254	1.6360
0,4	0.0032	0.0065	0.999	2.5557	0.7027	0.3521
RUN 2						
T	**Rsq**	**Fstat**	**pValue**	**DW**	**pValue**	**MSE**
0,05	0.1857	0.456	0.723	2.5712	0.7094	0.1670
0,1	0.4304	1.51	0.305	2.7916	0.9876	0.2526
0,15	0.4759	1.82	0.244	2.8626	0.8859	0.2807
0,2	0.5555	2.5	0.157	3.0750	0.5808	0.3284
0,25	0.5155	2.13	0.198	3.0494	0.6059	0.2509
0,4	0.1835	0.449	0.727	2.7181	0.8841	0.1352

Graphical analysis

From Figure 1 it can be seen that residuals seem to be normally distributed for cross-validation and prediction sets, so the assumption to use the linear regression are held.

From Figure 2 and Figure 3 can be also stipulated that for both datasets the regressions fit the data quite well. It does not fit well only the outliers that occur during the crisis (implied volatility peaks). On predictive dataset (Figures 4 and 5) it can be seen that regression line falls quite close to the data points. Therefore, there is an apparent potential of the model to predict well further volatility. This holds for both the 1st and the 3rd order models.

Let's analyze auto-correlation Durbin-Watson criteria:

From Tables 3 and 4 it can be seen that no autocorrelation tendencies are present as p-value of Durbin-Watson coefficient is mostly much larger than 0.05 significance level except a few maturities. So, it can be concluded that autocorrelation is not a problematic factor for current task.

Results for the 1st order regression model

Let's analyze R-squared, t- and F-statistics, MSE from the Table 3.

Firstly, let's take a look at the cross-validation set.

For RUN1 the value of r-squared falls approximately between 0.12 and 0.23 while for RUN2 it is between 0 and 0.4 and increases progressively for larger maturities. These values are not impressive, however r-squared cannot be used alone to make any meaningful conclusions. In terms of t- and F-statistics the predictor proves to be insignificant, i.e. it's p-value is around 0.2 in average which is superior than 0.05 significance level for both runs except some maturities for the predictive run.

The discrepancy between two runs is explained by ill-posedness of Dupire's model and the multiplicity of solutions, i.e. different inputs for different runs. Nevertheless, in terms of r-squared and even other statistics (the values are close to 0.2) such fit can be accepted if other measures prove model's significance.

From MSE it can be seen that the fit of the model on validation points is quite good as MSE values fall between 0.08 and 0.23 that means that no more than 23% of difference between the observed implied volatility and the value obtained from our model has been received. In general such difference means quite good prediction.

Secondly, let's take a look at the predictive set.

For both RUN1 and RUN2 the value of r-squared falls approximately between 0 and 0.14. This means quite low quality of the fit, however this can be explained by the presence of volatility jumps in this dataset as local volatility at the dates 01/17, 02/14 and 02/22 (see Table 1) is now a part of the fitted dataset (Table 2). In terms of t- and F-statistics the predictor proves to be definitely insignificant, i.e. it's p-value is much superior than 0.05 for both runs and lays between 0.15 and 0.98.

From MSE it can be seen that the fit of the model on validation points is quite good as MSE values fall between 0.11 and 0.24 that means that no more than 24% of difference between the observed implied volatility and the value obtained from our model has been received. In general such difference means quite good prediction.

So, for both cross-validation and predictive datasets the 1st order linear regression proves to predict well out-of-the-sample values of implied volatility. Despite some bad fit and insignificance of t- and F-statistics (caused by the presence of crisis outliers in the data) the model is able to predict well the implied volatility.

Let's see if the 3rd order model can improve overall results.

Results for the 3rd order regression model

Let's analyze R-squared, t- and F-statistics, MSE from the Table 4.

Firstly, let's take a look at the cross-validation set statistics.

For RUN1 the value of r-squared falls approximately between 0.39 and 0.7 while for RUN2 it is between 0.15 and 0.66. These values show that the fit is very good and outperforms significantly the 1st order regression.

Let's look at F-statistics. For RUN1 p-value is between 0.04 and 0.27 while for RUN2 it's between 0.06 and 0.79. For some maturities null hypothesis can be rejected and it can be said that the predictor is significant, however any pattern of significance depending on maturity for both runs can't be seen, so it can't be said that for one maturity the model is more significant than for another in general case. So, in terms of F-statistics the overall model proves to be insignificant.

MSE is between 0.08 and 0.22 for RUN1 and around 0.2 for RUN2. So, the local volatility can be considered as quite good predictor of the implied volatility on validation points. However, it can't be said that this model is better in comparison with the 1st order regression in terms of the predictive ability.

Secondly, let's take a look at the predictive set.

For RUN1 the value of r-squared falls approximately between 0 and 0.75 while for RUN2 it is between 0.18 and 0.55. These values show that the fit is very good and outperforms significantly the 1st order regression.

Let's look at F-statistics. For RUN1 p-value is between 0.02 and 0.99 while for RUN2 it's between 0.15 and 0.72. For some maturities null hypothesis can be rejected and it can be said that the predictor is significant, however any pattern of significance depending on maturity for both runs can't be seen, so it can't be said that for one maturity the model is more significant than for another in general case. Thus, in terms of F-statistics the overall model proves to be insignificant.

MSE is between 0.35 and 3.46 for RUN1 and around 0.28 for RUN2. So, the local volatility can't be considered as quite good predictor of the implied volatility on validation points. However, such observation of MSE pushes us to an interesting conclusion: if the 3rd order model doesn't predict well on out-of-the-sample dataset while the 1st order model does well with the same input, it's possibly because crisis jumps are included into the sample and basically overfit the model.

None of the models proves stable significance of local volatility as a predictor of the implied volatility but both can be actually used for practical purpose as they predict well out-of-the-sample implied volatilities.

So, on cross-validation dataset the 3rd order linear regression proves to predict well out-of-the-sample values of implied volatility as well as the 1st order model.

The advantage of the 3rd order model is that it fits well the data from the sample. The disadvantage is that sometimes overfitting can be observed that causes bad predictions, so in general it's actually safer to apply the 1st order model to datasets which contain the jumps.

Conclusions

In this work the 1st and the 3rd order polynomial linear regressions are used to see if at-the-money local volatility can predict the implied volatility and the results on out-of-the-sample validation points within two different datasets are validated.

Firstly, it was observed that residuals seem to be normally distributed for cross-validation and prediction sets, so the assumption to use the linear regression are held.

Secondly, auto-correlation using Durbin-Watson criteria was analyzed and concluded that no autocorrelation tendencies were present and it was stipulated that autocorrelation is not a problematic factor for current task.

Thirdly, it was concluded that neither the 1st nor the 3rd order model proves the significance of the local volatility as a predictor of the implied volatility but both can be actually used for practical purpose as they predict well out-of-the-sample implied volatilities. This is an important practical result as it means that complex non-linear relationship between implied and local volatilities formalized by Dupire can actually be reduced to simpler linear relationship that demonstrates reasonable discrepancies.

Fourth, and most importantly it was concluded that the 3rd order linear regression fits the sample data much better but for the reasons of overfitting in general it's actually safer to apply the 1st order model to datasets which contain the jumps.

References

1. Dupire B. Pricing with a smile. Risk, 7(1), 18-20 (1994).
2. Christensen, B.J., Prabhala, N.R. The relation between implied and realized volatility, Journal of Financial Economics, 50, 125—150 (1998).
3. Historical Options Data. URL: https://www.ivolatility.com/data/historical_data2.html
4. Mayhew S.: Implied volatility, Financial Analysts Journal, 51 (4), 8–20 (1995), doi: 10.2469/faj.v51.n4.1916
5. Cerf R. Asymptotic convergence of genetic algorithms. Advances in Applied Probability, 30(2), 521–550 (1998). URL: https://doi.org/10.1239/aap/1035228082
6. Bondarenko M., Bondarenko V. On dynamics of at-the-money local volatility calibrated from time series of VIX option. Modern engineering and innovative technologies, 12(6), 1-20 (2018). URL: http://dx.doi.org/10.21511/nfmte.7.2018.01
7. Wang Y. (2012). A Well-Posed Algorithm to Recover Implied Volatility. URL: https://www.diva-portal.org/smash/get/diva2:506716/FULLTEXT01.pdf
8. Derman E., Kani I., Zou J. (1996). The Local Volatility Surface: Unlocking the Information in Index Option Prices. Financial Analysts Journal - FINANC ANAL J. 52. 25-36. 10.2469/faj.v52.n4.2008.
9. Gatheral J., Hsu, E., Laurence P., Ouyang C., Wang T.H. Asymptotics of Implied Volatility in Local Volatility Models. Mathematical Finance. 22. 10.1111/j.1467-9965.2010.00472.x. (2009).
10. Ben Hamida S., Cont R. Recovering volatility from option prices by evolutionary optimization. Journal of computational finance, 8(4), 18-33 (2020). URL: http://dx.doi.org/10.21314/JCF.2005.130

Paper received : 09.03.2021 Printed : 01.06.2021