MODELS FOR OPTION PRICING BASED ON EMPIRICAL CHARACTERISTIC FUNCTION OF RETURNS

KAROL BINKOWSKI and ANDRZEJ KOZEK

Department of Statistics, Macquarie University
Sydney NSW 2109, Australia
E-mail: kbinkows@gmail.com, akozek@science.mq.edu.au

Abstract. The standard Merton-Black-Scholes formula for European Option pricing serves only as approximation to real values of options. More advanced extensions include applications of Lévy processes and are based on characteristic functions, which are more convenient to use than the corresponding probability distributions. We found one of the Lewis (2001) general theoretical formulae for option pricing based on characteristic functions particularly suitable for a statistical approach to option pricing. By replacing the unknown theoretical characteristic function with the empirical one the obtained model can be considered as a consistent estimator of the original Lewis formula. We explore the behaviour of this model on empirical data and conclude that it is necessary to allow for two additional implied parameters to obtain option pricing superior to other models reported in the literature.

1. Implied volatility as a cure for inadequacy of the MBS theory. The Merton-Black-Scholes (MBS) theory of option pricing (Black & Scholes (1973), Merton (1973)), based on the assumption of Gaussian distribution of logarithmic returns, results in the famous and well known formula for the value of a European Call option

\[ C_0(S, K, v, T, r) = S \cdot N(d_1) - K \cdot \exp(-rT)N(d_2), \]

where

\[ d_1 = \frac{\ln(S/K) + (r + v^2/2)T}{v\sqrt{T}}, \]

\[ d_2 = d_1 - v\sqrt{T}, \]

and where \( N(.) \) denotes the cumulative distribution function (cdf) of the standard nor-
mal distribution. The MBS formulae (1)-(3) reveal principal, but not all factors driving European Option prices: asset price $S$, strike price $K$, volatility (standard deviation) $v$ of the asset return, time to option expiration $T$ and risk-free interest rates $r$.

The MBS theory has been from the beginning considered, also by the authors of the theory, only as approximation to the real value of the option (cf. Field & Jaycobs (1992)). Deviations of the real option prices from the MBS model are caused by many factors, including the following ones

- the mood of the market is influenced by hopes and fears generated by news and resulting in violation of the assumption that logarithmic returns of asset prices follow normal distribution,
- real market is not perfectly efficient,
- liquidity is always limited, or,
- the conditional volatility of stock returns is not constant.

It has been quickly noticed that by adjusting volatility $v$ in the MBS formula one can numerically fit quite a large range of values for option prices. Hence, despite its drawbacks, the standard MBS formula plus a single empirically adjusted parameter $v$ are practically sufficient to describe the real option price. The adjusted volatility $\tilde{v}$ is called an implied volatility. Though the implied volatility is varying slower and in a much smaller range than the real option prices do, yet on real markets it changes over time and depends both on time to expiration and on strike. This resulted in a significant research aiming to improve our understanding of the nature of the implied volatility.


3. The MBS for Lévy processes via characteristic functions. Non-gaussian theories of option pricing do not allow simple explicit formulae like the MBS. Starting with Merton (1973) remark on his equation (39) on p. 167 and including more recent theories by Carr et al. (1998), Bakshi & Madan (2000) and Lewis (2001), it became clear that the MBS formulae and their extensions can be conveniently expressed using characteristic function of distributions of logarithmic returns. This observation leads to Fourier and Fast-Fourier Transforms (and also other transforms like in Borovkov & Novikov (2002)) techniques of option price evaluation. This is useful in cases where characteristic function is easier to calculate than the cdf of the returns. A summary of the approach based on characteristic functions can be found in Lewis (2001). One of the equivalent formulae derived in Lewis (2001) appears particularly convenient to our statistical approach.
4. The Lewis formula. Let us recall Carr et al. (1998) and Bakshi & Madan (2000) contribution in deriving for the Variance-Gamma (VG) process a valuation for an European Call option in a form similar to the classical MBS formula (1):

\[ C_1(S_0, K, \phi_T, T, r) = \Pi_1 S_0 - K e^{-rT} \Pi_2, \]

where

\[ \Pi_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \text{Re} \frac{\exp(-iu \log K) \phi_T(u-i)}{iu \phi_T(-i)} du \]

and

\[ \Pi_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \text{Re} \frac{\exp(-iu \log K) \phi_T(u)}{iu} du, \]

where \( \phi_T(u) \) is a characteristic function of the logarithm of the stock price process \( X_T = \log(S_T) \) at the expiration of the Call option.

Lewis (2001) extended formulae (4)-(6) onto general Lévy processes. In our project we work using his formula

\[ C_2(S_0, K, \phi_T, T, r) = S_0 - \sqrt{S_0 K} e^{-\frac{rT}{2}} e^{\frac{wT}{2}} \pi \int_0^\infty \text{Re} \left[ e^{-iu(\log S_0 K T + wT)} \phi_T(-u - \frac{i}{2}) \right] \frac{du}{u^2 + \frac{1}{4}}, \]

(cf. eq. (3.11) in Lewis (2001)), where the Mean Martingale Correcting Term (MMCT) \( w \) is determined by a martingale condition

\[ w = -\frac{\log(\phi_T(-i))}{T}. \]

Unlike the ARCH-type discrete time models or the Heston model, the formulae for option pricing based on characteristic functions do not model explicitly volatility, however they do so implicitly. The Lévy process behind the formulae can be interpreted as a Brownian motion with time flying at random speed and driven by a so called subordinating process. Researchers, including Carr et al. (2003), report excellent agreement of the corresponding models with the observed option prices, however it is fair to note that all parameters in these models are implied parameters.

5. Empirical and implied parameters. The Lewis formula (7) is particularly useful for a non-parametric statistical study where the characteristic function can be estimated based on historical data of daily asset returns. There is here no need to assume any particular family of infinitely divisible distributions like VG, Hyperbolic or Generalized Hyperbolic. Instead of estimating parameters of a chosen a priori particular subfamily of infinitely divisible distributions one can introduce a smaller number of implied parameters improving our understanding of how the real market prices the options.

In what follows we strictly distinguish between **empirical** and **implied** parameters.

**Empirical parameters** are estimators of the corresponding quantities based on historical data of stock returns.

**Implied parameters** are fitted (calibrated) using knowledge of the observed option prices and, possibly, also historical data of stock returns.
In the models considered in Section 7 our only empirical parameter is the empirical characteristic function estimated based on the most recent \( n = 120 \) daily returns.

6. Consistency of the Empirical Lewis Formula. In Binkowski & Kozek (2010) we extended results of Csörgő & Totik (1983) on convergence of empirical characteristic functions on increasing intervals onto convergence over rectangles increasing to strips in the complex domain parallel to the real axis. This extension implies that the empirical Lewis formula is converging to the original value when the sample size of returns grows to infinity, i.e. that it is a consistent estimator of the theoretical formula (7). Comparison of the behaviour of this estimator with real option prices presented below in Table 1 shows however far from excellent average option valuation. In contrast, other models, like in Carr & Madan (1999), with implied rather than estimated parameters, show excellent behaviour. This prompts us to allow up to two parameters in (10) to be chosen to fit the empirical option prices. In agreement with the introduced terminology, these parameters are referred to as implied parameters. The necessity of allowing implied parameters to be used along with the empirical ones, suggests that knowledge of the distribution of returns is not sufficient for getting high precision in recovering real option prices.

7. Five models with empirical and implied parameters. Our assumption that \( X_t = \log(S_t) \) is adequately modeled as a Lévy process implies that in decomposition

\[
X_T = X_0 + \sum_{j=1}^{p} (X_{j\Delta} - X_{(j-1)\Delta}),
\]

where

\[
T = p\Delta,
\]

the logarithmic returns \( R_j = (X_{j\Delta} - X_{(j-1)\Delta}) \) are i.i.d. random variables. Hence, and since \( X_0 \) is a known present value of \( \log(S_0) \), we get

\[
\phi_T(u) = (\phi_R(u))^p,
\]

where \( R \) has the same probability distribution as \( R_j \)'s. This allows us to use knowledge of probability distribution of logarithmic returns over a short period of time to infer the probability distribution of the logarithm of \( S_T \) at the option expiry. This is particularly useful working with historical data: we can estimate a characteristic function of \( \log(S_T) \) by applying (9) to empirical characteristic function (ECF) of logarithmic returns on short time intervals of length \( \Delta \).

We consider pricing of a European Call option, with time to maturity \( T \) and strike price \( K \), by replacing Lewis formula (7)–(8) with its empirical version

\[
\hat{C}_n(S_0, K, \hat{w}_n, T, r, p) = S_0 - \sqrt{S_0K} e^{-\frac{rT}{2} + \hat{w}T} \int_0^\infty \text{Re} \left[ e^{-iu(\log S_0 - rT + \hat{w}T)} \hat{\phi}_n \left( -u - \frac{i}{2} \right) \right] \frac{du}{u^2 + \frac{1}{4}},
\]

where the ECF \( \hat{\phi}_n(u) \) is based on \( n \) most recent daily logarithmic returns and \( p \) is the number of days to option expiration. Since \( T \) is expressed in years, \( \Delta \) is the fraction of the
year corresponding to one day. The presence of $p$ in (10) may look therefore redundant. However we will allow $p$ in some of the considered models (labeled below as Models 2 and 4, respectively) to be calibrated to real option prices and hence it is practical to accept this notation. Whenever '$p$' is considered as a parameter we use notation $\tilde{p}$, while symbol $p$ will be used to denote the number of days to option expiration.

Let us recall that $\hat{w}_n$ in (10) denotes the empirical version of the MMCT $w$ and is given by

$$\hat{w}_n = -\frac{\log(\hat{\phi}_n(-i))}{\Delta}. \quad (11)$$

In Models 3 and 4, we will replace $\hat{w}_n$ with a parameter $\tilde{w}$ which will be calibrated to option prices.

In our study we consider the following five models.

**Model 1:** no implied parameters. The parameter $\hat{w}$ is estimated and the theoretical $p$ equals the number of days to option expiration. In this model only historical data of $n = 120$ of the most recent returns of the asset (here the DAX index) are used.

In this model there are no implied parameters using additional information, like the real option prices and improving performance of the model.

**Model 2:** $\hat{w}$ is estimated and the implied $\tilde{p}$ is calibrated. In this model only one implied parameter $\tilde{p}$ is allowed. By fitting this parameter to historical option prices we in fact update the information available in the model about the most recent changes in the market. If the fitted $\tilde{p}$ is greater than the number of days to expiration $p$, then this suggests that the real market time flies slower. Clearly, $\tilde{p} < p$ indicates that the market time flies faster. Hence, this implied parameter is updating our model with the most recent information about the behaviour of the subordinating process.

**Model 3:** $\tilde{w}$ is implied and $p$ equals the number of days to option expiration. By allowing $\tilde{w}$ to be calibrated to the real option prices instead of using the empirical $\hat{w}$ we are introducing a kind of test on suitability of the risk neutral martingale measure introduced via Esscher transform. Since the considered market model is incomplete there exist many equivalent martingale measures (Perrakikis & Ryan (1984), Eberlein & Jacod (1997), Huang (2004)). However, by allowing $\tilde{w}$ to be calibrated without any restriction we do not have guarantee that $\tilde{w}$ corresponds to any of the possible risk neutral measures. Yet, it is safe to interpret $\tilde{w}$ as a parameter which can via calibration bring the valuation function $\hat{C}_n$ closer to real option prices.

**Model 4:** implied are both $\tilde{w}$ and $\tilde{p}$. We allow $\tilde{w}$ and $\tilde{p}$ to be jointly calibrated to real option prices. Following interpretation of these parameters given for Models 3 and 4, respectively, we allow both a small correction to the option price based on historical data (via $\tilde{w}$) and an update on the actual flow of the market time (via $\tilde{p}$). Our experiments with DAX index data suggest that this is the best model, i.e. leading to the best agreement with the real option prices.

**Model 5:** the VG model with implied 3 parameters of the VG probability distribution. This is the Carr & Madan (1999) model applied to our present data set. The VG distributions are parameterized with 3 parameters, which are calibrated to the real option prices.
Let us note that typically calibration and estimation result in different and incompatible set of parameters. The fact that in option pricing the calibration gives better results than estimation of the parameters suggests that historical data are not sufficient for precise option valuation. The market time may be changing at a different pace than the subordinator of the Lévy process implies, or, may be, as the market is incomplete, another equivalent martingale measure should have replaced the popular Esscher transform.

8. The underlying DAX index and ODAX options data. We tested performance of our models on historical data consisting of Deutsche Boerse AG DAX index (XETRA: GDAXI, ISIN: DE0008469008) and European Call Options ODAX (ISIN: DE0008469495) written on the index, and traded on Eurex. The options data include daily close price, strike price, and the time to maturity. The strike prices are set at 50 points intervals. The options have been recorded on Eurex exchange between the 1st of June 2006 and the 17th of May 2007 (243 business days). The number of maturities change over time and range from 1 to 6 of the closest ones to expiration. We used the data obtained from the Securities Industry Research Center of Asia-Pacific (SIRCA Ltd., http://www.sirca.org.au). Interest rates for this period have been taken from the European Central Bank web site (http://www.ecb.int/). From the options data set we chose only the most traded options, i.e. the 3 or 4 strikes nearest to the spot price. There were 2985 such options in our data set.

9. Precision of option pricing for Models 1-5. Denote by $\bar{C}(T, K)$ the observed option price (i.e. the daily settlement price) of an option with time $T$ to expiration and strike $K$. We used the following option pricing errors to rank the considered five models. The Mean Absolute Error (MAE):

$$\frac{1}{N} \sum_{l=1}^{N} |\bar{C}(T_l, K_l) - \hat{C}_n(S_l, K_l, \hat{w}_{n,l}, T_l, r_l, p_l)|,$$

the relative MAE:

$$\frac{1}{N} \sum_{l=1}^{N} |\bar{C}(T_l, K_l) - \hat{C}_n(S_l, K_l, \hat{w}_{n,l}, T_l, r_l, p_l)|/|\bar{C}(T_l, K_l)|,$$

the Root Mean Square Error (RMSE):

$$\sqrt{\frac{1}{N} \sum_{l=1}^{N} (\bar{C}(T_l, K_l) - \hat{C}_n(S_l, K_l, \hat{w}_{n,l}, T_l, r_l, p_l))^2},$$

the relative RMSE:

$$\sqrt{\frac{1}{N} \sum_{l=1}^{N} [(\bar{C}(T_l, K_l) - \hat{C}_n(S_l, K_l, \hat{w}_{n,l}, T_l, r_l, p_l))/\bar{C}(T_l, K_l)]^2},$$

where $N$ is the number of option prices depending on the considered case.

---

1International Securities Identification Number.
In Figure 1 we present results of option pricing using models 1-4, for the 1st of June 2006. We used \( n = 120 \) prior days to calculate the ECF, i.e. we take \( \hat{\phi}_{120}(u) \) and \( \hat{w}_{120} \) given by (11). The spot price was \( S_0 = 5707.59 \), the interest rate was \( r = 0.035 \). We used four maturities, corresponding in Figure 1 to the four slightly declining horizontal layers of four points, respectively, and with 4 strikes for each maturity, corresponding to vertical columns, respectively. Values of strikes are presented on the horizontal axes and prices of options are on the vertical axes. The black line is the option payoff. Blue rectangles denote the real ODAX option prices and red circles denote the model prices. We considered here the following maturities: 18, 53, 109 and 200 days.

In Table 1 we present average errors taken over all cases considered for one year of option pricing. Model 1, with model parameters specified completely by the past 120 returns of the DAX index, performs rather poorly, suggesting that additional information is needed to improve option pricing quality. Taking into account how little information was provided to the model, only the underlying returns over past 120 days, it would be however unfair to disregard this model. This experiment rather shows importance of the calibration combined with the historical behaviour of the underlying asset.

In Models 2–5, historical option prices have been used for models calibration. Model 2, with calibrated parameter \( \tilde{p} \) performs much better than Model 1, even better than Model 3 with the calibrated parameter \( \tilde{w} \). This suggests that it is very important to
model properly the subordinator and the speed of the market time flow using the most recent information available via option prices used in calibration. Another correction, in Model 3, coming via the calibrated $\tilde{w}$, is also significant when compared with Model 1. It may either suggest usage of an inappropriate martingale transformation, or $\tilde{w}$ can be considered as another parameter leading to an improved fit to the empirical option prices. Model 4, using the most recent 120 DAX returns and adjustments $\tilde{p}$ and $\tilde{w}$ calibrated to option prices performs best. Table 1 shows that Model 4 with 2 implied parameters outperforms the other four, including Model 5 with 3 implied parameters, considered by Carr & Madan (1999).

**Table 1.** Measurement of errors for Models 1-5 for one year of pricing options.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>36.7306</td>
<td>2.8594</td>
<td>5.2871</td>
<td>0.3306</td>
<td>2.3868</td>
</tr>
<tr>
<td>Relative MAE</td>
<td>0.3908</td>
<td>0.0883</td>
<td>0.1172</td>
<td>0.0055</td>
<td>0.0156</td>
</tr>
<tr>
<td>RMSE</td>
<td>44.9672</td>
<td>4.0897</td>
<td>8.7297</td>
<td>0.4680</td>
<td>3.3061</td>
</tr>
<tr>
<td>Relative RMSE</td>
<td>1.2103</td>
<td>0.9575</td>
<td>0.9893</td>
<td>0.0269</td>
<td>0.0307</td>
</tr>
</tbody>
</table>

Figure 2 shows results of calibration for Model 2. Each point with coordinates $(p, \tilde{p})$ refers to one maturity: $p$ denotes the true number of days to option expiration while $\tilde{p}$ denotes the value obtained from calibration. There are 774 calibrated $\tilde{p}$. The straight line represents regression $\tilde{p} = \alpha_0 + \alpha_1 p$, with coefficients $\alpha_0 = 3.1510$ and $\alpha_1 = 0.8866$. It shows a linear relationship between those $p$ that are true and those that are implied, however the figure shows heteroscedasticity of the data and outliers.

Figure 3 shows results of calibration of parameter $\tilde{w}_n$ in Model 3. It contains a plot of $\tilde{w}$ versus $\tilde{w}_n$. Each point represents one maturity, altogether the 774 calibrated $\tilde{w}$. The straight line represents regression $\tilde{w} = \alpha_0 + \alpha_1 \tilde{w}_n$, with coefficients $\alpha_0 = 0.1590$ and $\alpha_1 = 1.1417$. Like in Figure 2 we observe heteroscedasticity in the data and outliers, as well as skewness.

We are not presenting similar plots for Model 4, for which both parameters are calibrated jointly. The relation between the empirical MMCT, the number of days to expiration and the implied parameters, is not so straightforward.

Figure 4 presents how the empirical $\tilde{w}_{120}$ (the black bold line), the maximum (red line) and the minimum (blue line) of the implied $\tilde{w}$ from Model 3 change over the year. The maximum and minimum are taken over different implied parameters related to different maturities for each day, respectively. The empirical MMCT has been used in Models 1 and 2.

Similarly, Figure 5 shows the empirical $\tilde{w}_{120}$ (black bold line), and the maximum (red line) and the minimum (blue line) of the implied $\tilde{w}$ from Model 4. The maximum and minimum are taken over different implied parameters related to different maturities for each day, respectively.

Figure 6 shows close prices of DAX between 120 working days prior to the 1st June 2006 (marked by a vertical red line) and the 17th of May 2007.
Fig. 2. Number of days to expiration $p$ vs. implied $\tilde{p}$, based on Model 2 for one year of pricing options.

Fig. 3. Empirical $\hat{w}_n$ vs. implied $\tilde{w}$, based on Model 3 for one year of pricing options.
Fig. 4. Empirical $\hat{w}_{120}(u)$ based on 120 historical log-returns between the 1st of June 2006 and the 17th of May 2007 for each day (black line) and the implied minimum (blue line) and maximum (red line) $\tilde{w}$ for Model 3.

Fig. 5. Empirical $\hat{w}_{120}(u)$ calculated from 120 historical log-returns between the 1st of June 2006 and 17th of May 2007 for each day (black line) and the minimum (blue line) and maximum (red line) of the implied parameter $\tilde{w}$ for Model 4.
Fig. 6. Close prices of DAX between 120 working days prior to the 1st June 2006, and 17th May 2007 (blue line). Red line indicates the 1st June 2006.

Fig. 7. The 12 series of the implied $\tilde{p}_n$, obtained for Model 2 between the 1st of June 2006 and the 17th of May 2007.
10. Behaviour of the implied parameters. In Figures 7 and 8 we show behaviour of the implied parameters $\tilde{p}_n$ and $\tilde{w}_n$ as time series. They can be modeled using regression methods with stationary errors providing further insight into the option pricing process and even presenting a commercial value to Market Makers. Indeed, tools allowing to use, at least as a starting point at the opening of the day, the implied parameters for today estimated based on the data available till yesterday, can be quite valuable allowing to avoid miss-pricing of options by those who are responsible for guarantying the liquidity of the market.

To fit the regression-time-series models we used a statistical package EViews. We fit series of implied $\tilde{p}_n$ and $\tilde{w}_n$ to the following model (cf. Quantitative-Micro-Software (2007), Chapter 26)

$$y_m = x'_m \beta + u_m,$$  \hspace{1cm} (12)

$$u_m = \alpha_1 u_{m-1} + \alpha_2 u_{m-2} + \epsilon_m,$$  \hspace{1cm} (13)

where $\beta' = [\beta_1, \beta_2]$ are regression parameters, $\alpha_1, \alpha_2$ are parameters of a hidden autoregressive (AR) model driving the regression noise $u_m$ and $x_t$ is a vector of explanatory variables. Let us note that model (12)-(13) can be also presented equivalently without the hidden AR component in the following way

$$y_m = x'_m \beta + \alpha_1 (y_{m-1} - x'_{m-1} \beta) + \alpha_2 (y_{m-2} - x'_{m-2} \beta) + \epsilon_m,$$

however, representation (12)-(13) allows a clear interpretation. We take $y_m$ to be either $\tilde{p}_n$ or $\tilde{w}_n$, respectively. In the case of implied $\tilde{p}_n$ we include the number of days to option

Fig. 8. The 12 series of implied $\tilde{w}_n$, obtained for Model 3 between the 1st of June 2006 and the 17th of May 2007.
expiration as an explanatory variable, and in the case of \( \tilde{w}_n \) we include the empirical MMCT as an explanatory variable.

Let us note that since \( p = T/\Delta \) the implied parameter \( \tilde{p}_n \) depends on the time to option expiry

\[
\tilde{p} = \frac{T}{\tilde{\Delta}}.
\]

Hence, in the case of an ideal model, where \( \tilde{\Delta} \) is constant the \( \tilde{p}_n \) is a linear function of the time to the expiry \( T \). This justifies our use of \( p \) as a regressor in the time-series modeling. We have observed that values of obtained implied parameters \( \tilde{w}_n \) are near values of the empirical MMCT. This suggests that in modeling series of implied \( \tilde{w}_n \) we should regress on the empirical MMCT.

We refer to Binkowski (2008) for a detailed analysis of the considered regression-time-series models.

11. Conclusions. We combined the modern theory of European options pricing allowing the logarithms of returns of the underlying asset to be driven by a fairly general Lévy process with nonparametric estimation of the characteristic function. A clear distinguishing between estimated and implied parameters leads to a better understanding of the role of the historical data for the underlying asset and access to the historical option prices in fitting and calibrating of the model.

By choosing for calibration two implied parameters we obtained an option pricing model, labeled in the paper as Model 4, giving excellent agreement with the real options prices. Let us note that for every maturity the implied parameters have been the same for all strikes and allow modeling of their behaviour over time by regression methods with stationary errors. Standard forecasting methods allow to predict the future values of the implied parameters reducing risk of miss-pricing e.g. by market makers.

References


