INTRADAY STOCK RETURN DISTRIBUTION FOR BLACK-SCHOLES OPTION PRICING

Daniel W. Tsang and Tak David Cheung
City University of New York-Queensborough Community College, Bayside, New York 11364

Keywords: Levy-Gaussian convergence, volatility, investment strategy, diffusive system, bootstrap

Abstract
The availability of intraday stock/index return in the web facilitates the improvement of return volatility estimation over the traditional method that is based on inter-day return data. Truncated Levy process distribution is used to extract the intraday return distribution parameters. The calibration to the volatility for Black-Scholes option pricing is studied using the data from Levy-Gaussian convergence in physical diffusive systems as well as using the empirical implied volatility values. It appears that intraday return distribution parameters give short-term call option prices closer to the market values. Robustness is investigated using bootstrap data. The resulting volatility variability is studied in contrast to the observed intraday fluctuation of the implied volatility. Application to investment strategy is also discussed.

Introduction
The volatility parameter in the Black-Scholes (B-S) Option pricing model has been a subject of intense research ever since Black and Scholes introduced it in 1973 [1]. The original volatility of the underlying asset was estimated using end of day prices. The return follows a Gaussian distribution with its standard deviation as the volatility value. As web information emerges in the nineties, volatility estimation based on 5-minute interval prices also becomes a reality. The CBOE implied volatility VIX data is a good example. Similar advances in research have also been reported throughout the open literature and financial engineering has emerged as a new discipline in the last ten years.

The 5-minute interval S&P500 index futures dataset provides solid evidence that the stock return deviates much from a Gaussian distribution[2]. The dataset spans several years from 1991-1995 with some 75,000 data points. With success in describing the recent advances of physical science diffusion systems, Levy distribution has also become a favorite tool for the analysis of the short time interval stock dataset. This paper addresses the application of truncated Levy distribution to intraday dataset with a few hundreds data points. The distribution width is calibrated to the volatility value of the B-S model. Computer generated bootstrap data is also used to study robustness. The QQQ stock and OEX index are used as application examples.

Project Design
Louis Bachelier, a student of Henri Poincare, published his thesis "Theory of Speculation" within Gaussian diffusion framework in 1900[3]. Since then physical science diffusion reveals many examples of Non-Gaussian diffusion in recent studies of turbulent phenomena[4]. Transport variables such as temperature fluctuation in turbulent cell study, force distribution in plastic materials, light and microwave fluctuation in turbid media obey a characteristic function proportional to exp-((constant)^k |variable|^k) for $0 < k < 2$, which is commonly referred to as Levy distribution. The case $k = 2$ is the usual Gaussian probability density function after performing the Fourier transform.

For both stock return and turbulent transport, the strength of Levy distribution lies in its ability to account for statistical outliers. Outliers are exactly what the stock market has and the return is seldom Gaussian especially in a short term period. Recently, Matacz published the use of the geometric truncated Levy process for the stock price and a hedging strategy minimizing some appropriate measure of risk because the risk-free B-S delta strategy is not available[5]. The report uses the $k = 1.2$ Levy distribution and it also shows the theoretical conditions that multiple truncated Levy processes converge to an overall Gaussian process, which has already been observed experimentally in the physical science such as photon diffusion with small absorption[6].

Wilmott presents the trinomial random walk model and deduces that the stock price probability satisfies the Kolmogorov equation with Taylor series approximation[7]. This means that the B-S option value is equivalent to the present value of the expected payoff at expiry under a risk neutral random walk for the underlying stock. If S is the stock price, r is the risk-free date, σ is the volatility, dX is a Brownian/Wiener process, then $dS = rSdt + \sigma$

SdX. The up or down stock movement is drawn from a Gaussian N(0,1) distribution and the zero movement probability is equal to (1- the sum of the probabilities of up and down). This is the origin of the Gaussian walker. We propose that the Levy walker takes on a different distribution dX'. From the Levy case k=1 the walker process dX' can be drawn from a Cauchy distribution. This is accomplished by taking the movement to be tan (θ) with θ as uniformly distributed because then tan (θ) follows a Cauchy distribution. Appropriate truncation is needed to control the divergence of tan (θ) at 90 degrees. Both the trinomial and Levy-Cauchy random walk model do not have B-S risk-free delta hedging strategies. Nonetheless, this Levy-Cauchy walker may be important in analyzing short-term un-hedged position. In the long run, this Levy-Cauchy walker becomes a Gaussian walker if one imposes a truncation on the Cauchy tail. As mentioned above, θ should be set at an appropriate value to reflect the practical truncation limit. The study of the truncated Levy-Cauchy walker would forecast the Gaussian parameters. Particle movement with multiple scatterings in thick turbid medium gives rise to Gaussian diffusion. Einstein first described it in the context of mean free path length λ. The displacement square $<R^2> \sim (\lambda^2/\tau)*t$ with τ being the average time between collision. The Gaussian standard deviation scales as sqrt(time). Note that τ depends on the particle speed. For photon transport in a thin sample (thickness β) with a single scattering center, the scattering transverse coordinate $\beta\tan(\theta)$ can be written as $\Lambda\sin(\theta)$, with Λ as the scattering path length. When β increases to the maximum thickness where Cauchy profile is still observed, $\Lambda = \beta / \cos(\theta_t)$ with θ_t being the truncation angle given by $\tan^{-1}$(extreme profile data point/β) and Λ becomes the mean free path λ (about 10% error). The maximum Cauchy β forecasts the Gaussian λ with time τ.

Among the Levy distributions, the rationale for working with the Cauchy distribution, with characteristic function exp-((constant)^k |variable|^k) for k = 1, is two folded. The Fourier transform for the probability density function gives the Cauchy distribution for k = 1 and the Gaussian distribution for k = 2. The random walk requirement is that the sum of N steps should have the same distribution as a single step. The general Levy class satisfies the requirement. However only the Gaussian and Cauchy have explicit analytical expressions. All other Levy cases require series expansions and are not ready to be used in data curve fitting. Analysis based on probability of return to zero P(0) for the Levy series has been proposed with no statistical justification. In this paper a statistical approach is chosen instead. An observed Gaussian random walker in long time scale actually behaves like a Cauchy random walker in some appropriate small time scale, with truncated Cauchy tail for a practical system. It is well known in early quantum mechanics that x-ray scattering form factor is Cauchy-like, which is generally called the Lorentzian in physical science and engineering. There are always small physical dissipation effects that effectively truncate the long tail. The requirement of thin scattering sample thickness is equivalent to small time scale in stock movement. Second, the fat-tail Levy distribution has been an active research subject in option pricing ever since Mandelbrot started using it for cotton price dynamics in 1963[8]. This subject is also consistent with the observation that equity option volatility smile implied distribution has fatter tails as compared to the usual lognormal distribution[9]. As discussed earlier, Matacz found that k = 1.2 works well for the Australian All Ordinaries Index database. The database is divided into different time interval datasets. Each dataset's peak return probability value is collected to form a decay curve for the extraction of k. We believe that the value of 1.2 is an artifact related to the use of 30-min time unit in the database. We postulate that the k = 1.0 parameter fit is a better fit for intraday fluctuation with smaller time interval. Our data shows that the intraday distribution obeys the Cauchy scaling property (to be discussed later).

The qualm of imposing exact symmetry property of the Cauchy distribution on economics data could be resolved using the psychology principle that investors are rational. A trivial algorithm is proposed. The up/down portion of the distribution represents those investors who believe in a positive/negative return from that asset. The profitable strategy is to buy call/put option if the asset return is positive/negative. In principle the call option calculation should use a theoretical symmetrical distribution generated by projecting the positive return data onto the negative domain. The opposite algorithm should be performed for put option calculation. In practice the collected return data (OEX, QQQ) has one peak only and is quite symmetrical within 10% for dataset up to 5-min time interval. In other words, the market is efficient and rational statistically when observed with a large number of data points. The few percent asymmetry property is probably real. It may be used to investigate the arbitrage assumption in today's web technology environment but is certainly outside the scope of this paper. Furthermore given the bid-ask spread of the traded option, this symmetry qualm/solution remains in the theoretical domain.

The use of Bayesian estimation is necessary in principle because the variance of Cauchy distribution,

$\text{Prob}(x_k | \alpha, \beta) \sim \beta / (\pi(\beta^2 + (x_k - \alpha)^2))$,

diverges and the central limit theorem does not apply. Sample variance estimate drawn from the distribution would not converge to a single number as the sample gets larger, but would keep increasing without limit[10]. This issue diminishes gradually as truncation takes over the divergence in practical application for daily operation. In Bayesian estimation, the choice of prior usually depends on the parameter[11]. We found that uniform prior works well in this project. After the Bayesian estimation, the truncated Levy distribution is imposed so as to forecast the Gaussian volatility for the B-S model. The β value forecasts the standard deviation σ value of the Gaussian walker which is valid in longer time duration. The σ value is then used as the time dependent volatility and is inserted into the B-S equation, with a short expiry and with the usual B-S assumption of negligible transaction cost etc[7]. The α value is the Levy-Cauchy drift and could be used to replace the risk-free interest rate in the B-S equation for the analysis of un-hedged short-term position[7,].

Results and Discussions

An example of the S&P 100 index (OEX) return data is shown and Bayesian estimation is used to get the Levy-Cauchy parameters. QQQ return data (not shown) behaves similarly.

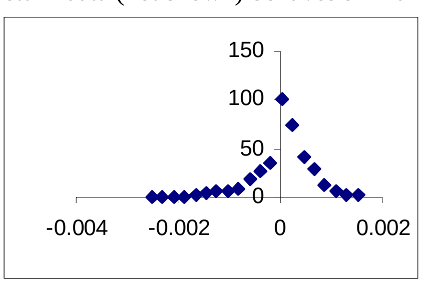

Fig 1A: 4/22/02 OEX return data based on the intraday data as posted by wallstreetcity.com. The y-axis is the normalized frequency and the x-axis is the return. Bayesian method gives $\beta = 0.00025$ using the positive return data.

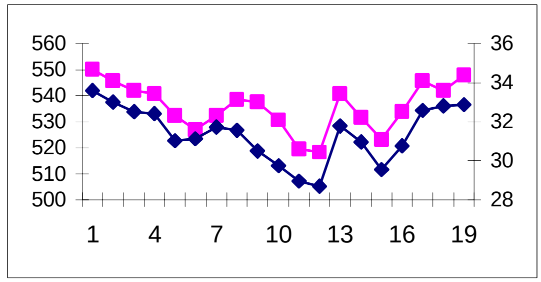

Fig 1B: OEX (blue/lower) and QQQ (pink/upper) closing price data versus trading days starting from 4/22/02 to 5/16/02. The left y-axis is OEX price and the right y-axis is QQQ price. The data is available from wallstreetcity.com historical data section.

The intraday price data is available every 90 seconds from wallstreetcity.com. Historical data quotes are also available (Fig 1B). Sampling time of longer duration can be constructed from the 90-sec intraday data. The convolution of two Cauchys is a new Cauchy with width equals to $\beta_1 + \beta_2$. Linear dependence of the full width at half maximum (FWHM) 2β is observed up to 6-min time interval dataset for both OEX and QQQ. We use a 10% error criterion to accept linear dependence because a Gaussian FWHM is 2.35σ. The maximum volatility value is then scaled up for input in the B-S model (252 trading days) that allows risk-free hedging strategy. The 6-min β and truncation values yield a mean free path λ. The λ value is the volatility σ with a corresponding τ for OEX. The calculation of τ from first principle is not a trivial task because it depends very much on the particle/stock speed. Therefore the τ value for this project is determined by calibrating σ to the observed flat implied volatility at the beginning 3 days (average) of a 30-day call contract. This τ value would be used for the remaining days on the contract. The τ for OEX is about 120 min. The τ for QQQ is about 350 min. Whether the difference is related to the fact that there are 100 stocks in OEX and QQQ is a single stock may be pure speculation at this moment. It is noted that

Motorola also has similar τ value as QQQ and the QQQ associated index NDX has τ ~180 min in our investigation (to be published elsewhere). This means that OEX would achieve Gaussian behavior in 8 days if the same observed intraday activity persists also for 8 days while QQQ/Motorola would achieve it in about 25 days. In contrast Matacz reported that the Australian index takes 19 days to cross over to Gaussian regime. In summary, the β and truncation values are used to track the variability during the 30-day contract period. The interest rate is taken as the 3-month T-bill rate. It is interesting to note that the above 6-min Levy period is also consistent with the 5-min frequency in ARCH time series analysis of S&P 100 index[12].

The Levy walker is the precursor of the Gaussian walker. This estimation of future volatility is better for predicting the return in the near future than to use the typical last 30-day stock closing values. The Levy extraction provides a quantitative measure of the B-S values in the same relevant time frame as the fluctuation of the underlying stock price, which is reflected in the immediate intraday volatility. The resulting B-S values are closer to the market values.

Call contract example I

OEX May-02 Call contract with strike price at $540. The data is from etrade.com. The risk-free interest rate is from 3-month T-bill rate.

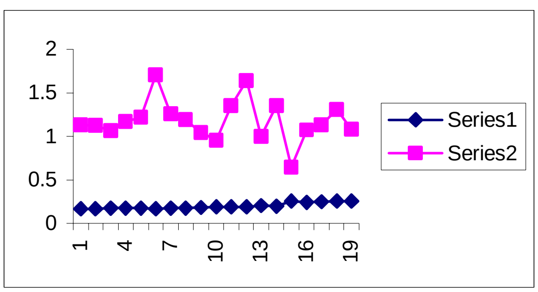

Fig 2A: The upper curve is a plot of OEX (trading price/B-S price) versus trading days. The data starts on 4/22/02 and ends at 5/16/02. The lower curve is the 30-day volatility. The data is posted by us.etrade.com at 5 pm.

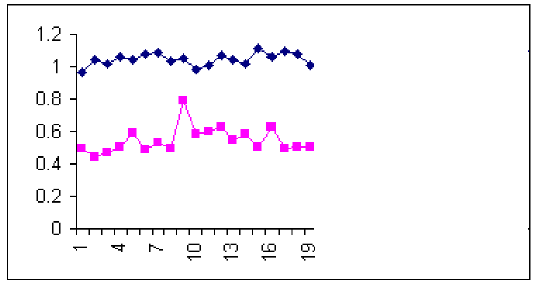

Fig 2B: The lower curve is 2*(volatility) based on Levy extraction versus trading days as defined in Fig 2A. The upper curve is a plot of (trading price/B-S price based on the Levy volatility values) versus trading days.

The trading price is calibrated to the B-S price such that the effect of stock price, interest rate, and expiry duration are eliminated. The remaining significant parameter is the volatility. The deviation ratio of trading price to B-S price is used to illustrate the inaccuracy of the B-S calculation using the 30-day volatility. The price movement is the driving force that affects the volatility and increases the call trading price (Fig 1B). The local price minimum is on 5/7/02. The small τ enables the deviation ratio (Fig 2A) to follow the price movement (Fig 1B). If the volatility values are correctly estimated, the deviation ratio should be relatively flat. In other words, the observation of the deviation ratio fluctuation means that the B-S volatility values are wrong. The traditional volatility estimation performs worse during large intraday price movements. There is a significant time lag between the 30-day volatility and deviation trends as depicted in Fig 2A. On the other hand, when the B-S price is based on the volatility extracted by the Levy method, the time lag is eliminated and the deviation ratio values are close to 1 because now the B-S prices are closer to the trading values. Also the Levy extracted volatility in Fig 2B closely follows the deviation ratio values (upper curve) in Fig 2A as a function of time. Note that the extracted volatility values in fig. 2B are enlarged by a factor of 2 to illustrate the calculation noise. The mean free path error is amplified during the scaling in sqrt(t) given the small τ. The fluctuation is within the bid-ask spread and the high-low price of the day. Therefore the Gaussian volatility in the B-S calculation should be obtained from the Levy extraction rather than the usual 30-day extraction.

Call contract example II
Nasdaq100Trust (QQQ) May-02 Call contract with strike price at $30 (no dividend). The data is from etrade.com. The risk-free interest rate is from 3-month T-bill rate.

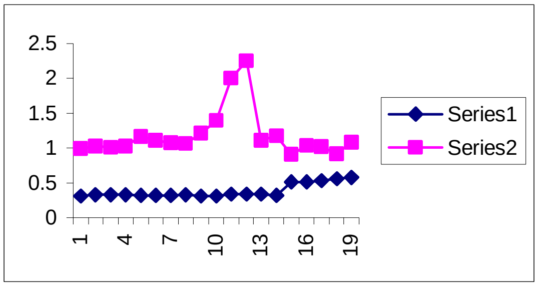

Fig 3A: The upper curve is a plot of QQQ (trading price/B-S price) versus trading days. The data starts on 4/22/02 and ends at 5/16/02. The lower curve is the 30-day volatility. The data source is same as in Fig 2A.

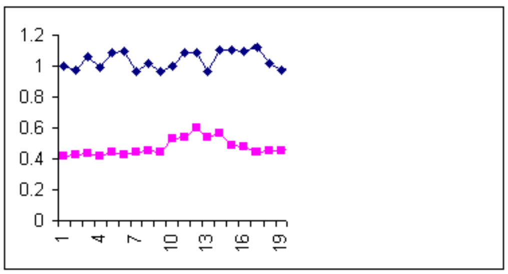

Fig 3B: The lower curve is volatility based on Levy extraction versus trading days as defined in Fig 3A. The upper curve is a plot of (trading price/B-S price based on the Levy volatility values) versus trading days.

QQQ also has a time lag between the 30-day volatility and deviation trends as depicted in Fig 3A. Similarly, when the B-S price is based on the volatility extracted by the Levy method, the time lag is eliminated and the deviation ratio values are close to 1 because now the B-S prices are closer to the trading values. Also the Levy extracted volatility in Fig 3B closely follows the deviation ratio values (upper curve) in Fig 3A as a function of time. Note that the QQQ deviation ratio trend differs slightly from the price trend. This may be due to the large QQQ τ value.

New hedging delta values are readily obtainable from the N (d_1) formula[7]. The demonstration of an OEX or QQQ hedging portfolio requires much simulation and is beyond the scope of this project. Nevertheless, judging from the above graphs, the new delta values are closer to the implied volatility delta values for both OEX and QQQ since the Levy values are close to the implied volatilities.

The above discussion could also be applied to contracts with different strike prices. Intraday data from 9:30 am to 3 pm could be used to compute the Levy values. There is still an hour to do market actions.

The success of using intraday return distribution also may be used to explain the observed volatility smile variation across strike prices. The intraday stock price usually moves within a limited range and the volatility value thus extracted is only applicable at/near the money. For out/deep in the money, discrepancy such as the volatility smile does not signal the break down of the B-S framework, but rather the need of an alternative volatility algorithm.

Bootstrap of intraday return data
The 90-sec intraday data is used as the original dataset to generate bootstrap data using the standard block resample technique for time series data. The stock data: S1, S2, S3, S4, ... are organized into blocks of length 4, 6-min block, such as Z1 = (S1, S2, S3, S4), Z2 = (S5, S6, S7, S8)…. Block re-sampling for bootstrap data is performed to generate new data sequence, for example, Z100, Z1, Z38,.... …. and bootstrap sample sizes = 199 is used.

The 5/7/02 dataset is chosen here for discussion purpose. This particular dataset is chosen because both the OEX and QQQ prices movements did not cross over to new strike prices on that day. This condition is necessary for comparison because the corresponding option based volatility indices VIX/QQV are computed from at/near the money positions. The results show that bootstrap Cauchy width b* has standard deviation values within 20% of the avg (b*) for OEX and QQQ returns.

Other block lengths are also tested. For 9-min blocks, Cauchy width b* has standard deviation values within 5% of the avg (b*). For 15-min blocks, the bootstrap Cauchy width b* shows standard deviation values within 2% of the avg (b*). On the other hand, the 5/7/02 observed standard deviation for VIX/QQV is about 0.35/0.9 with a mean of about 22/40. This translated to 1.6%/2.3% for OEX/QQV respectively and agrees well with the bootstrap based on 9-min block re-sampling. Detailed future study could be performed to reveal the optimal bootstrap block length. From an empirical viewpoint, the variability of the extracted volatility for OEX/QQQ returns based on bootstrap data is shown to be consistent with the intraday fluctuation of the implied volatility VIX/QQV index. This intraday distribution methodology would provide addition signal for market bottom/support level and thus investment opportunity.

Conclusion

This paper explores the use of Levy statistics for intraday return volatility. The calibration to the Gaussian volatility is performed using Levy-Gaussian convergence observed in physical science diffusion. Although this approach is in mark contrast to the ARCH modeling that used intraday 5-min stock data and implied volatility data, it does arrive at the conclusion that 6-min is the optimum study period in high frequency intraday data. However, there were different approaches in the open literature on how to handle the overnight trading effect. The ARCH dataset spans over months and overnight trading effect was modeled as a single transaction. Similarly the S&P 500 dataset mentioned in the introduction section ignores the jumps between data subsets. This overnight trading effect is not an issue in our current model. This project, for the first time as far as we know, focuses on the time evolution of an individual call contract that is correlated to the intraday volatility; and thus eliminates the time lag effect of using the traditional 30-day volatility in price calculation. It offers an alternative approach to measure the volatility of a stock/index and the result is consistent with the option based volatiles indices VIX and QQV via the bootstrap investigation. The extension to put option application should be similar. The method retains the B-S scheme and thus the popular delta hedging strategy.

Acknowledgement: We are thankful to the questions and discussions in the conference. Eric Cheung provided some data support. DT serves in the Business Department. TDC serves in the Physics Department and also as Doctoral Faculty, CUNY-Graduate Center

References

1. (a) Fisher Black and Myron Scholes " The pricing of Options and Corporate Liabilities: Journal of Political Economy, 81, pp637-654, 1973. (b) Robert Merton, " Theory of Rational Option Pricing" Bell Journal of Economics and Management Science 28, pp141-183, 1973

2. Rama Cont, "Scaling and correlation in financial time series" European Physical Journal B, cond-mat/9705075, 1997

3. Louis Bachelier, "Theorie de la Speculation" (1900), English translation in "The Random Character of Stock Market Prices" by Paul H. Cootner,ed., MIT Press, Cambridge pp17-78, 1964

4. (a) Joseph Klafter, Michael Shlesinger and Gert Zumofen "Beyond Brownian Motion" Physics Today, vol 49, pp33-39, 1996. (b) 9. Leo Kadanoff "Turbulent Heat Flow: Structures and Scaling", Physics Today, vol 54, 34-39, 2001

5. Andrew Matacz, " Financial Modelling and Option Theory With The Truncated Levy Process", International Journal of Theoretical and Applied Finance, vol 3, no 1, pp143-160, 2000.

6. E. Chan, T. Cheung, and P. Wong, "Statistical Modeling of Photon Diffusion in Turbid Media with Application to Bioimaging", Proceedings of 1996 Annunal Meeting of American Statistical Association, Physical and Engineering Sciences Section, pp187-192, 1996.

7. Paul Wilmott, "Derivatives: The theory and practice of financial engineering", Wiley, 1998.

8. B.B. Mandelbrot "The fractal geometry of nature" Freeman Press, New York 1982 and J Business, vol 36 pp394, 1963.

9..John Hull, "Options, Futures, and Other Derivatives", Prentice Hall, 2000

10. Jeffrey H. Dorfman "Bayesian Economics Through Numerical Methods", Springer, 1997

11. Ronald Mahieu and Rob Bauer, "A Bayesian analysis of stock return volatility and trading volume", Applied Financial Economics, vol 8, pp671-687, 1998.

12.(a) T.G. Andersen & T. Bollerslev, "Answering the skeptics: yes standard volatility models do provide accurate forecasts", International Economic Review 39, pp885-905, 1998. (b)Bevan Blair, Ser-Huang Poon, and Stephen Taylor, "Forecasting S&P 100 Volatility: The incremental Information Content of Implied Volatilies and High Frequency Index Returns", online working paper, Lancaster University, November, 2000.