Options and Jump Diffusion Models

Jacob Aylmer, Harshil Cherukuri, Dhruv Kanchi, Om Mehta, William Parente

May 1st, 2024

1. Abstract

Created by Robert Merton in 1976, jump diffusion models have emerged as a powerful method of financial modeling. These models help capture the fashion in which the market reacts to breaking news, or sudden, discontinuous movements in asset prices. When unexpected incoming news is processed by the market, volatility of stock prices increases mightily. This rapid change is not perfectly represented by Brownian Motion in a standard diffusion model, so incorporating the component of a jump is imperative for accurately modeling stock and option prices. Traditionally, in the aftermath of the publication of Merton's first paper, papers on jump diffusion models present new prospective models that researchers believe can effectively measure jumps. Our paper offers a new perspective, in that we have aggregated a few different models to compare and contrast the results from each of them. We have created a simple detection model, the Lee-Mykland Model, and a machine learning based isolation forest model, each of which has a different strategy to define and capture a jump. It is important to highlight that each model has its own strengths and weaknesses, and by standardizing our results and comparing models to each other, we are able to come up with a comprehensive analysis of the financial market from January 2010 - December 2011 in terms of movement of both SPY equity and its respective call and put options. Our three models produce significant results in terms of calculating the number of total jumps, as well as categorizing each one as an up jump or a down jump, as well as computing the lambda for each model. We have been able to identify the most severe jumps throughout our time series data, as well as detect anomalies through an unsupervised learning model. Overall, this paper provides a sound perspective into the expanding landscape of jump diffusion models in finance, with massive implications for financial analysts.

2. Introduction

In an effort to more accurately model the abnormal variation of stock prices, Dr. Robert C Merton pioneered a new revolutionary model called "Jump-Diffusion". This is considered more accurate than the previously accepted diffusion model, because it incorporates the jump component. The jump is incorporated into the model in order to capture the way that the market reacts to breaking news. When incoming news is learned by market participants, stock prices become much more volatile and proceed to rise and fall much faster than during normal times. This rapid change is not represented well by a normal diffusion model, so the jump component is useful to more accurately model stock prices. Using this more accurate model, we can analyze how option prices are affected when jumps occur in the underlying stock. We plan to explore this relationship in depth, and provide visual representations such as histograms to show the impact of jumps in stock prices on the prices of their options. Based on a variety of different factors such as time of the occurrence of the jump, size of the jump, and volume of the jump, we will create a clustering technique to categorize the jumps.

3. Literature Review

Stock prices move for two reasons: normal vibrations and abnormal vibrations. Normal vibrations are modeled by geometric Brownian motion, while abnormal vibrations result from new information causing significant changes in stock prices, modeled by a jump process (Merton 1976). Merton's influential paper introduced a formula for pricing options based on underlying stock price and time to maturity, forming the basis for subsequent research in jump diffusion.

Building on Merton's work, we plan to utilize a double-exponential jump-diffusion model developed by Kou (2002) to calibrate stock price movements using intraday data.

The concept of jump diffusion processes, introduced in the paper by Merton (1976), extended the Black-Scholes-Merton model by incorporating discontinuous stock returns (Feng and Linetsky 2008). This paper outlined ideal conditions for continuous trading, emphasizing the martingale property in stock price dynamics and introducing the formula for pricing options based on stock price and time to maturity. It highlighted the increased value of options on stocks with jump components, profoundly influencing subsequent research in finance.

Feng and Linetsky (2008) proposed an innovative method for valuing options within jump-diffusion models, using partial integro-differential equations to improve precision and efficiency in financial engineering. Their paper also analyzed the impact of time steps on model accuracy and compared their method with alternative valuation techniques developed by researchers like Broadie-Yamamoto and Merton.

Addressing limitations of the Black-Scholes model, particularly its failure to account for the leptokurtic nature of stock returns and the volatility smile observed in option markets, a jump-diffusion model was introduced. This model incorporates a jump component to capture sudden changes in stock prices, offering a more accurate representation for pricing a variety of options (Kou 2002). The formula for Kou's double-exponential jump-diffusion model for stock price at time t is as follows:

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma)t + \sigma W(t)} \prod_{i=1}^{N(t)} V_i$$

Where W(t) is a standard Brownian motion, N(t) is a Poisson process with rate λ, and V_i is a sequence of independent identically distributed non-negative random variables such that Y=log(V) has an asymmetric double exponential distribution. μ is the drift and σ is the volatility, both are assumed to be constant.

Comparing various option pricing models, including the Black-Scholes model and the Merton Jump approach, Kremer and Roenfeldt (1993) evaluate their efficacy in pricing European call and put options using real-market data, providing insights into the practical and theoretical underpinnings of contemporary option pricing mechanisms.

Investigating the joint time series of the S&P 500 index and short-dated option prices, Pan (2002) reveals the presence of jump-risk premia, which respond swiftly to changes in market volatility. These premia play a crucial role in explaining volatility smirks observed in options data, highlighting the importance of understanding jump processes in refining financial models.

Our research focuses on the work on Merton and Kou, but the other papers were essential in acquiring a better understanding of modeling and analyzing jumps in stock price over time.

4. Methodology

4.1 Theory

We will begin by specifying the Jump-Diffusion model based on the framework established by Dr. Robert C Merton in his 1976 paper, "Option Pricing when Underlying Stock Returns are Discontinuous." We have defined the key components of the model, including the diffusion process representing normal stock price movements and the jump process capturing abrupt, discontinuous changes in stock prices. We utilized historical data from the Hanlon Lab to estimate the parameters of the Jump-Diffusion model. This involved estimating different aspects of our model, such as the diffusion coefficient, jump intensity, jump size distribution, and other relevant factors. We implemented a variety of statistical methods to optimize the model parameters based on the observed historical stock and option price data. We then took steps to clean and pre-process the obtained 15-minute data for SPY and its options from January 2010 to December 2011. This included handling missing data, outliers, and ensuring the data was in a suitable format for analysis. We were careful to aggregate the data appropriately, considering the required time intervals for modeling the Jump-Diffusion process effectively, as well as developing detection techniques to categorize jumps based on a variety of factors such as the time of occurrence, size, and volume of the jump. This involved applying algorithms to identify patterns and jumps in stock price. Methods such as k-means clustering or hierarchical clustering could have merit in helping us determine the most appropriate approach for categorizing jumps in a prospective extension of this project.

When it came to the implementation of the Jump-Diffusion model for analysis, we had to incorporate the results of our jump detection methods to understand how different types of jumps impact option pricing differently. Calculation of option prices based on the Jump-Diffusion model for different scenarios, considering the presence of jumps and their characteristics. A good way we demonstrated this was by generating visual representations to illustrate the impact of jumps in stock prices on the prices of associated options. The model's accuracy and predictive power was another area of analysis for us, considering its ability to capture the dynamics of option pricing during periods of stock price jumps. By following this methodology, we have provided a comprehensive analysis of the relationship between stock price jumps and option prices, utilizing the Jump-Diffusion model and incorporating insights from Dr. Robert C Merton's groundbreaking work in option pricing, as well as those who have contributed and developed his work after him.

4.2 Models

4.2.1 Jump Detection

Simple Detection Method We employed a simple detection method to set a baseline and gain a better understanding of the data we were working with. Under the simple detection method, a jump is identified when the return of the stock price exceeds the mean return by more than 2 standard deviations. Although stock returns more closely resemble a leptokurtic distribution, if we assume a normal distribution, this should identify slightly less than 5 percent of observations as jumps. This simple model can be a good point of reference when analyzing our other models.

Lee-Mykland Model As another way to identify jumps in a stock price, we utilize the Lee-Mykland model. This model uses a test statistic, which is the result of dividing the log return of a single period by a rolling standard deviation. This test statistic, along with two constants: C_n and S_n, are used to determine if the test is significant at level α. By using the rolling standard deviation as well as the calibrated constants, the Lee-Mykland model can be used to more accurately identify jumps in a stock price process. The formulas for the test statistic and constants are displayed below:

Definition 1. *The statistic $\mathcal{L}(i)$, which tests at time t_i whether there was a jump from t_{i-1} to t_i, is defined as*

$$\mathcal{L}(i) \equiv \frac{\log S(t_i)/S(t_{i-1})}{\widehat{\sigma(t_i)}}, \tag{7}$$

where

$$\widehat{\sigma(t_i)}^2 \equiv \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} |\log S(t_j)/S(t_{j-1})||\log S(t_{j-1})/S(t_{j-2})|. \tag{8}$$

$$C_n = \frac{(2logn)^{0.5}}{c} - \frac{log(\pi) - log(logn)}{2c(2logn)^{0.5}}$$

$$S_n = \frac{1}{c(logn)^{0.5}}$$

$$c = \sqrt{\frac{2}{\pi}}$$

If

$$\frac{|L(i)| - C_n}{S_n} > B^*$$

where

$$B^* = -log(-log(1 - \alpha))$$

then the test is significant and a jump is identified.

Machine Learning Isolation Forest Model Our third and final model aimed to construct a Jump Diffusion Model (JDM) using a machine learning approach, specifically leveraging the Isolation Forest algorithm. This powerful unsupervised anomaly detection technique excels in isolating outliers within datasets, capitalizing on the principle that anomalies are more susceptible to isolation than normal data points. By employing random decision trees and isolating instances in fewer splits, anomalies are rapidly identified, offering efficient outlier detection without relying on labeled data. This method proves particularly beneficial in scenarios where labeled datasets are scarce or impractical to obtain. Moreover, the Isolation Forest algorithm's robustness in handling high-dimensional and large datasets enhances its versatility across various domains. In our implementation, we utilized the Isolation Forest model to capture a total of 972 jumps within the dataset, with 487 classified as up jumps and 485 as down jumps. Notably, the mean percentage changes for up jumps and down jumps were 0.16219 and -0.38639, respectively. Through meticulous hyper parameter tuning, we optimized the Isolation Forest model's performance, achieving a lambda value of 0.04823 ($\lambda = 972/20,150 = 0.04823$). This approach allowed for a more unsupervised or semi-supervised learning model, given the challenges in measuring jumps across the entire dataset and the complexity of acquiring additional data.

4.2.2 Pricing

Using the parameters obtained from our jump detection methods, we implemented an options pricing model. Our model simulates random stock price paths with both geometric Brownian motion and jump aspects with discrete time steps. At each time step, there is a geometric Brownian motion portion that always slightly moves either up or down. Additionally, there is also the chance for a jump to occur, which is determined by a Poisson process, N(t), with lambda obtained from the Lee-Mykland detection method. If there is a jump at a time step, the stock price is multiplied by V=exp(Y), where Y follows a Laplace distribution with location (mean) and scale (standard deviation) calibrated from the jumps detected by Lee-Mykland. The formula for simulated stock price is displayed below:

$$S(t) = S(0)e^{(\mu-\frac{1}{2}\sigma)t+\sigma W(t)} \prod_{i=1}^{N(t)} V_i$$

After simulating 10000 stock price paths, we find the payoff of each path, and take the average to get the the theoretical price of the option. This is then compared to both the Black-Scholes price and the market price to evaluate performance.

5. Data Description

In our project, we used SPY and SPY Options Time Series Data from 2010 and 2011 using 15-minute interval prices to implement the Jump-Diffusion Model.

A RIC (Refinitiv Instrument Code) is an identifying code that gives us information about the option, including the underlying asset, the strike price, the maturity date, and whether it is a call or put option. As a result, we had to parse the RIC column for all of our data in order to extract the necessary information from the RIC.

During the data cleaning process, we strategically divided the dataset into two distinct frames based on our observations within the RIC column. One frame exclusively contained information related to the SPY ETF, while the other encompassed details pertaining to various SPY option chains within the specified time period.

Subsequently, we optimized the size of these data frames by eliminating unnecessary columns and filtering out entries associated with options that were not traded. Employing these straightforward data cleaning techniques resulted in the SPY data frame comprising 20,150 rows and 8 columns. Likewise, the SPY Options data frame consisted of 65,478,041 rows and 12 columns after the cleaning process.

Head of SPY Data:

	#RIC	Date-Time	Open	High	Low	Last	Volume	No. Trades
0	SPY	2010-01-04T08:00:00.000000000-05	112.12	112.180000	111.440000	112.140	1765282.0	428.0
1	SPY	2010-01-04T08:15:00.000000000-05	112.14	112.160000	112.090000	112.150	682644.0	674.0
2	SPY	2010-01-04T08:30:00.000000000-05	112.16	112.230000	112.140000	112.220	988952.0	873.0
3	SPY	2010-01-04T08:45:00.000000000-05	112.22	112.330000	112.220000	112.320	378532.0	1167.0
4	SPY	2010-01-04T09:00:00.000000000-05	112.33	112.460000	112.330000	112.390	554833.0	1641.0
5	SPY	2010-01-04T09:15:00.000000000-05	112.39	112.440000	111.440000	112.360	684084.0	2020.0
6	SPY	2010-01-04T09:30:00.000000000-05	112.37	112.750000	112.330000	112.730	12623404.0	31193.0
7	SPY	2010-01-04T09:45:00.000000000-05	112.73	112.940002	111.440000	112.830	9225048.0	25754.0
8	SPY	2010-01-04T10:00:00.000000000-05	112.83	112.960000	112.730000	112.900	8984505.0	20777.0
9	SPY	2010-01-04T10:15:00.000000000-05	112.91	113.100000	112.839996	113.085	4809410.0	13115.0

Tail of SPY Data:

	#RIC	Date-Time	Open	High	Low	Last	Volume	No. Trades
20141	SPY	2011-12-30T15:30:00.000000000-05	125.95	126.0200	125.889999	126.00	5011278.0	11412.0
20142	SPY	2011-12-30T15:45:00.000000000-05	125.94	126.1966	125.589996	125.60	14204786.0	26201.0
20143	SPY	2011-12-30T16:00:00.000000000-05	125.60	126.1966	125.470000	125.74	15233941.0	9353.0
20144	SPY	2011-12-30T16:15:00.000000000-05	125.71	126.1966	125.600000	125.66	1674857.0	726.0
20145	SPY	2011-12-30T16:30:00.000000000-05	125.66	125.6700	125.590000	125.63	137327.0	263.0
20146	SPY	2011-12-30T16:45:00.000000000-05	125.63	125.6400	125.450000	125.50	173197.0	420.0
20147	SPY	2011-12-30T17:00:00.000000000-05	125.51	125.6300	125.490000	125.60	104651.0	193.0
20148	SPY	2011-12-30T17:15:00.000000000-05	125.60	125.6500	125.600000	125.63	28035.0	63.0
20149	SPY	2011-12-30T17:30:00.000000000-05	125.64	125.6400	125.580000	125.58	9734.0	19.0
20150	SPY	2011-12-30T17:45:00.000000000-05	125.60	125.6300	125.600000	125.62	1300.0	4.0

Head of SPY Options Data:

	#RIC	Date-Time	Open	High	Low	Last	Volume	No. Trades	Root	Strike	Expiry	Option Type
0	SPYA161010000.U	2010-01-04T09:30:00.000000000-05	12.50	12.75	12.50	12.75	19.0	3.0	SPY	100.0	2010-01-16	call
1	SPYA161010200.U	2010-01-04T09:30:00.000000000-05	10.84	10.84	10.84	10.84	40.0	1.0	SPY	102.0	2010-01-16	call
2	SPYA161010400.U	2010-01-04T09:30:00.000000000-05	8.48	8.48	8.47	8.47	2.0	2.0	SPY	104.0	2010-01-16	call
3	SPYA161010500.U	2010-01-04T09:30:00.000000000-05	7.65	7.88	7.65	7.88	84.0	10.0	SPY	105.0	2010-01-16	call
4	SPYA161010600.U	2010-01-04T09:30:00.000000000-05	6.80	6.89	6.80	6.83	13.0	4.0	SPY	106.0	2010-01-16	call
5	SPYA161010800.U	2010-01-04T09:30:00.000000000-05	4.82	5.01	4.81	5.00	23.0	7.0	SPY	108.0	2010-01-16	call
6	SPYA161010900.U	2010-01-04T09:30:00.000000000-05	3.80	4.15	3.80	4.15	22.0	8.0	SPY	109.0	2010-01-16	call
7	SPYA161011000.U	2010-01-04T09:30:00.000000000-05	3.00	3.34	3.00	3.30	318.0	23.0	SPY	110.0	2010-01-16	call
8	SPYA161011100.U	2010-01-04T09:30:00.000000000-05	2.38	2.55	2.29	2.51	3458.0	99.0	SPY	111.0	2010-01-16	call
9	SPYA161011200.U	2010-01-04T09:30:00.000000000-05	1.71	1.83	1.56	1.81	7475.0	204.0	SPY	112.0	2010-01-16	call

Tail of SPY Options Data:

	#RIC	Date-Time	Open	High	Low	Last	Volume	No. Trades	Root	Strike	Expiry	Option Type
2587547	SPYX301112500.U	2011-12-30T16:00:00.000000000-05	0.02	0.02	0.01	0.01	1974.0	66.0	SPY	125.0	2011-12-30	put
2587548	SPYX301112600.U	2011-12-30T16:00:00.000000000-05	0.40	0.50	0.15	0.31	11729.0	272.0	SPY	126.0	2011-12-30	put
2587549	SPYX301112700.U	2011-12-30T16:00:00.000000000-05	1.40	1.41	1.26	1.27	350.0	23.0	SPY	127.0	2011-12-30	put
2587550	SPYX301112800.U	2011-12-30T16:00:00.000000000-05	2.45	2.45	2.27	2.27	185.0	11.0	SPY	128.0	2011-12-30	put
2587551	SPYX301113000.U	2011-12-30T16:00:00.000000000-05	4.39	4.39	4.39	4.39	5.0	1.0	SPY	130.0	2011-12-30	put
2587552	SPYX311213000.U	2011-12-30T16:00:00.000000000-05	15.10	15.10	15.10	15.10	10.0	1.0	SPY	130.0	2012-12-31	put
2587553	SPYA211212900.U	2011-12-30T16:15:00.000000000-05	0.92	0.92	0.92	0.92	3500.0	2.0	SPY	129.0	2012-01-21	call
2587554	SPYM211211800.U	2011-12-30T16:15:00.000000000-05	0.56	0.56	0.56	0.56	6500.0	2.0	SPY	118.0	2012-01-21	put
2587555	SPYM211212500.U	2011-12-30T16:15:00.000000000-05	2.11	2.11	2.11	2.11	6500.0	2.0	SPY	125.0	2012-01-21	put
2587556	SPYX301112700.U	2011-12-30T16:15:00.000000000-05	1.25	1.25	1.25	1.25	3500.0	2.0	SPY	127.0	2011-12-30	put

Augmented Dickey Fuller Test on SPY Returns:

```
        Augmented Dickey-Fuller Test

data:  returns
Dickey-Fuller = -32.164, Lag order = 27, p-value = 0.01
alternative hypothesis: stationary
```

6. Analysis

6.1 Jump Detection

Method	Total Jumps	Up Jumps	Mean Up Jump	Down Jumps	Mean Down Jump
Simple Detection	776	382	0.747%	394	-0.726%
Lee-Mykland	378	172	0.869%	206	-0.797%
Isolation Forest	972	487	0.162%	485	-0.386%

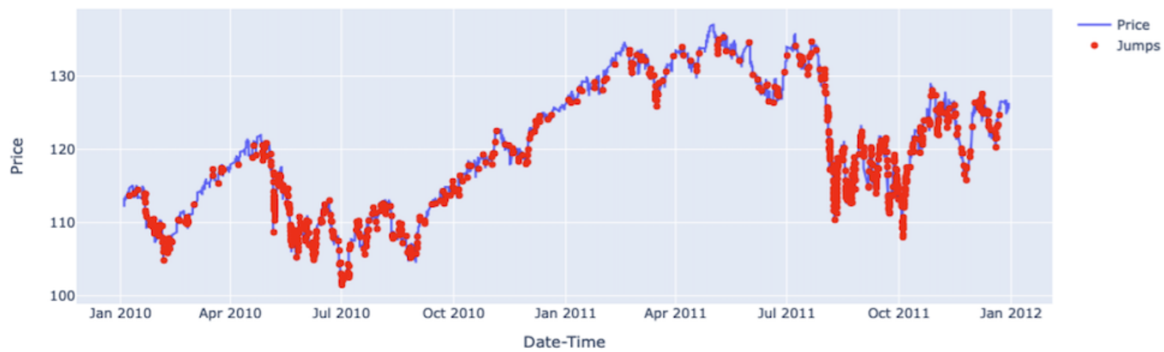
We observed some varying results from the two models we implemented. Under the simple model, we identified 776 jumps, over twice as many as with the Lee-Mykland model, which identified only 378. For the simple method, the percentage of observations that are classified as jumps is 3.85 percent, while the Lee-Mykland identified 1.88 percent as jumps. In support of the left skewed leptokurtic distribution that a stock return distribution should follow, there were more down jumps identified than up jumps when using both methods. Additionally, the absolute value of the jump size was larger for the Lee-Mykland jumps, which is intuitive because there were less jumps identified, meaning a more extreme movement was necessary. The stock price over time, along with the jumps identified by each model, are displayed further below.

6.2 Pricing

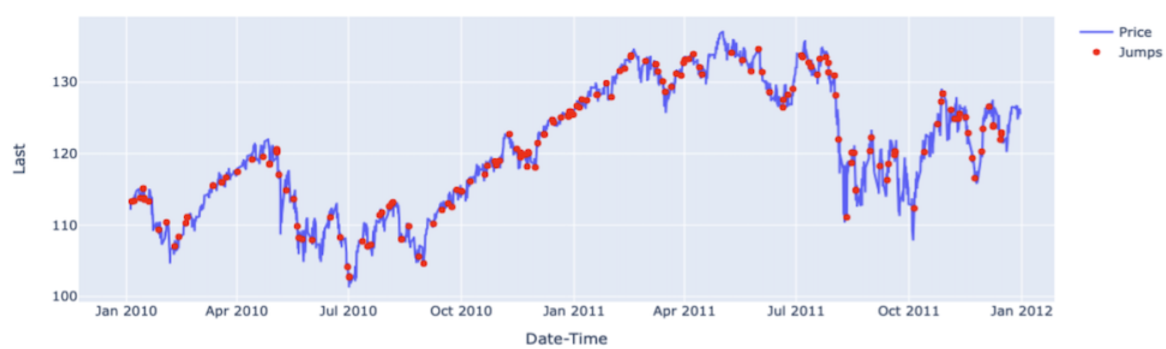
Model	January 10, 2010	June 15, 2010	October 5, 2010	February 25, 2011	August 8, 2011	November 14, 2011	Average
Jump Diffusion Model	0.3258	0.2679	0.1575	0.4286	0.3411	0.2854	0.3011
Black Scholes Model	0.3566	0.2732	0.1423	0.5093	0.3684	0.2788	0.3214

Using the parameters obtained from the detection methods above, specifically the Lee-Mykland model, we were able to successfully implement an options pricing model. We were able to use the total number of jumps divided by the total number of observations as an estimate for lambda (frequency) and the average size of the jumps as an estimate for intensity. Through our model, we were able to price options, both calls and puts, on SPY during 2010 and 2011. When comparing our model to the traditional Black-Scholes model, we ended up pricing options closer to their market price. On average, Black-Scholes was around 32% away from the market price. Meanwhile, our pricing model which incorporates jumps in stock price, was about 30% away from the market price. The 2% improvement is quite significant, especially in the world of options pricing.

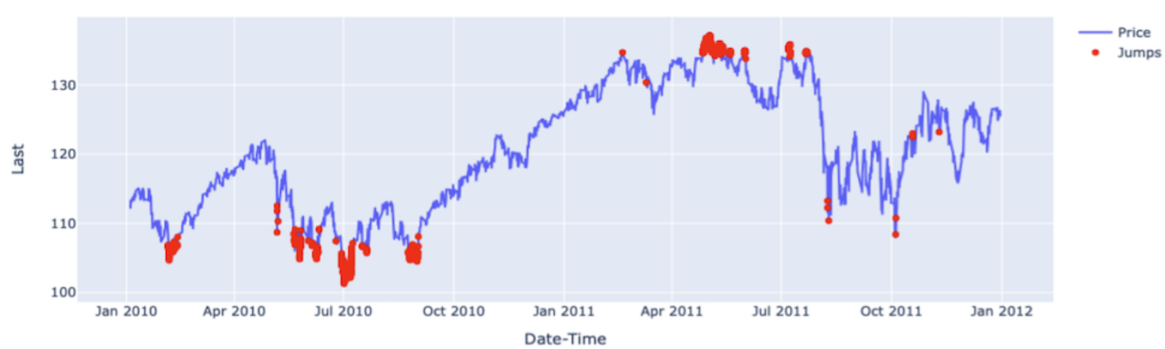
Simple Detection Model:



Lee Mykland Model:



Isolation Forest Model:



7. Conclusion and Next Steps

We are satisfied with the results we have obtained. We have been able to utilize a simple detection method as well as the Lee-Mykland and isolation forest models to identify jumps in stock price. All three methods produced promising results that we have utilized with an options pricing model to determine the impact of jumps on option prices, with Lee-Mykland providing the best results.

We further implemented an option pricing model which incorporates jumps in stock price. These moments of increased volatility are not accounted for in Black-Scholes, but are a major aspect in determining stock price movement. It is important to model the jumps in stock price in order to more accurately price options, and we were able to successfully do so. Our work resulted in a more accurate pricing model than Black-Scholes, which we are very pleased with.

We would like to acknowledge Professor Khaldoun Khashanah and Siyuan Ma for their support throughout the duration of this project. Their guidance proved to be invaluable and we would not have been nearly as successfully without their help.

8. References

Merton, Robert C. 1976. "Option Pricing When Underlying Stock Returns Are Discontinuous." Journal of Financial Economics 3 (1-2): 125–44. https://doi.org/10.1016/0304-405x(76)90022-2.

Feng, Liming, and Vadim Linetsky. "Pricing Options in Jump-Diffusion Models: An Extrapolation Approach." Operations Research 56, no. 2 (2008): 304–25. http://www.jstor.org/stable/25147189.

Kou, S. G. "A Jump-Diffusion Model for Option Pricing." Management Science 48, no. 8 (2002): 1086–1101. http://www.jstor.org/stable/822677.

Kremer, Joseph W., and Rodney L. Roenfeldt. 1993. "Warrant Pricing: Jump-Diffusion vs. Black-Scholes." The Journal of Financial and Quantitative Analysis 28 (2): 255. https://doi.org/10.2307/2331289.

Pan, Jun. 2002. "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study." Journal of Financial Economics 63 (1): 3–50. https://doi.org/10.1016/s0304-405x(01)00088-5.

Rydin Gorjão, L., Witthaut, D., and Lind, P. G. (2023). jumpdiff: A Python Library for Statistical Inference of Jump-Diffusion Processes in Observational or Experimental Data Sets. Journal of Statistical Software, 105(4), 1–22. https://doi.org/10.18637/jss.v105.i04

Mari, Carlo, and Emiliano Mari. "Gaussian clustering and jump-diffusion models of electricity prices: a deep learning analysis." Decisions in Economics and Finance 44, no. 2 (2021): 1039-1062.

Liu, Xiaobai, Qian Xu, Thuan Chau, Yadong Mu, Lei Zhu, and Shuicheng Yan. "Revisiting jump-diffusion process for visual tracking: a reinforcement learning approach." IEEE Transactions on Circuits and Systems for Video Technology 29, no. 8 (2018): 2431-2441.

Campbell, Andrew, William Harvey, Christian Weilbach, Valentin De Bortoli, Thomas Rainforth, and

Arnaud Doucet. "Trans-Dimensional Generative Modeling via Jump Diffusion Models." Advances in Neural Information Processing Systems 36 (2024).

Ramezani, Cyrus A., and Yong Zeng. "Maximum likelihood estimation of the double exponential jump-diffusion process." Annals of Finance 3 (2007): 487-507.

Villaplana, Pablo. "Pricing power derivatives: A two-factor jump-diffusion approach." Available at SSRN 493943 (2003).