Insights of Jumps in Small-Cap Stocks Proposal

My research idea lies mainly in the field of Stochastic Process, Statistics, Data Science and may a little bit Finance (because finance only serves as the background of my topic, which is about jump return). My research topic is highly original because stock jumps in big cap stocks have been well studied by Merton (1976)'s pricing model, Poison Process, or even more simple, like GBM (Geometric Brownian motion). However, while big cap stocks' jumps are characterized as their apparent cause and the fact that they do not self-excite, small cap stocks' jumps are more volatile because a slight change in trading volume or big-ask spread will cause the small cap stock to jump. So they constantly jump, some mathematician even assume that small-cap stocks' growth are completely consist of stock jumps, but martingale, of course.

My research's main goal is to find out the stochastic process and the distribution of the time interval of jumps in small cap stocks. Using this, we can approximate the probability of a jump happening at any given time of one stock's evolution. If possible, I can even, using statistics distribution model and stochastic process, find the expected jump size of that time (if there are one). I need to determine these variables because the most fundamental theory has fix parameters and the main task of my model is to build up a time dependent variable matrix.

$$dS_t = (\mu - \lambda\kappa)S_{t-}dt + \sigma S_{t-}dW_t + S_{t-}dJ_t$$

Where:

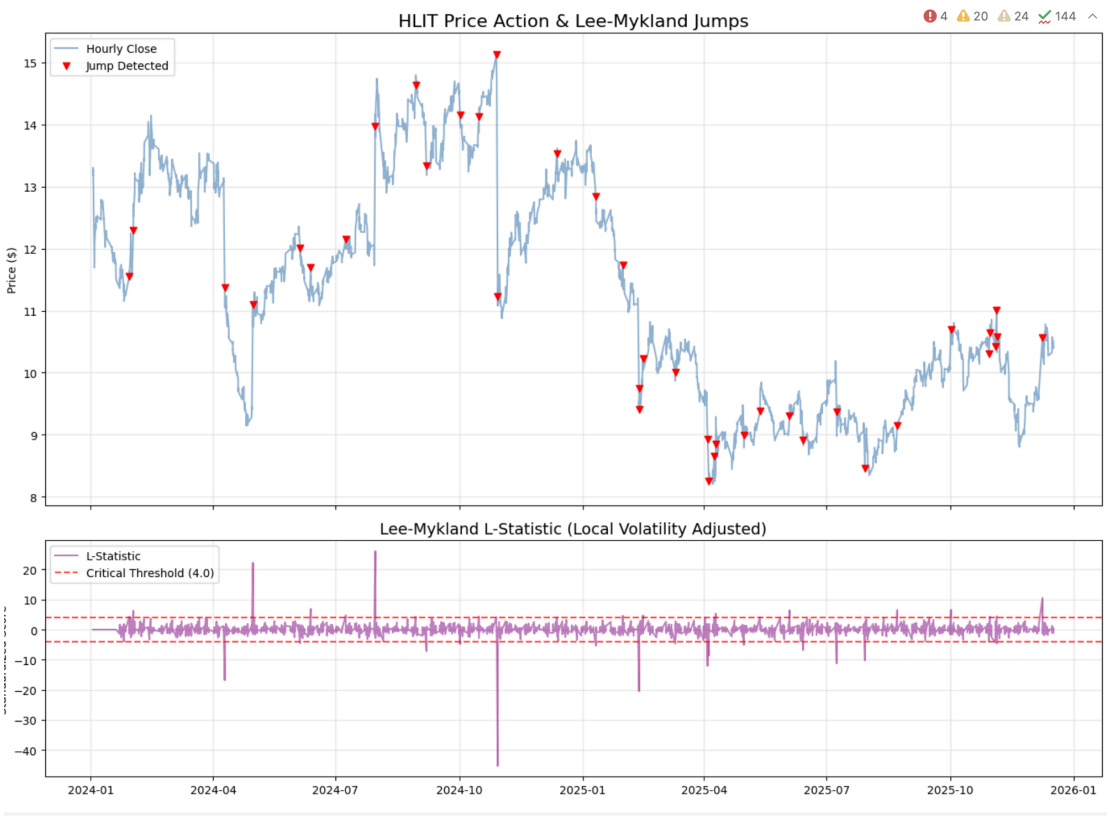
- W_t: Standard Brownian motion.
- N_t: A Poisson process with intensity λ.
- Y_i: The random jump size of the i-th jump.
- $J_t = \sum_{i=1}^{N_t}(Y_i - 1)$: The jump process.
- $\kappa = E[Y - 1]$: The expected relative jump size.

$$S_t = S_0 \exp\left(\left(\mu - \lambda\kappa - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right)\prod_{i=1}^{N_t} Y_i$$

$$\frac{dS_t}{S_{t-}} = (r - \lambda\kappa)dt + \sigma dW_t + d\left(\sum_{i=1}^{N_t}(Y_i - 1)\right)$$

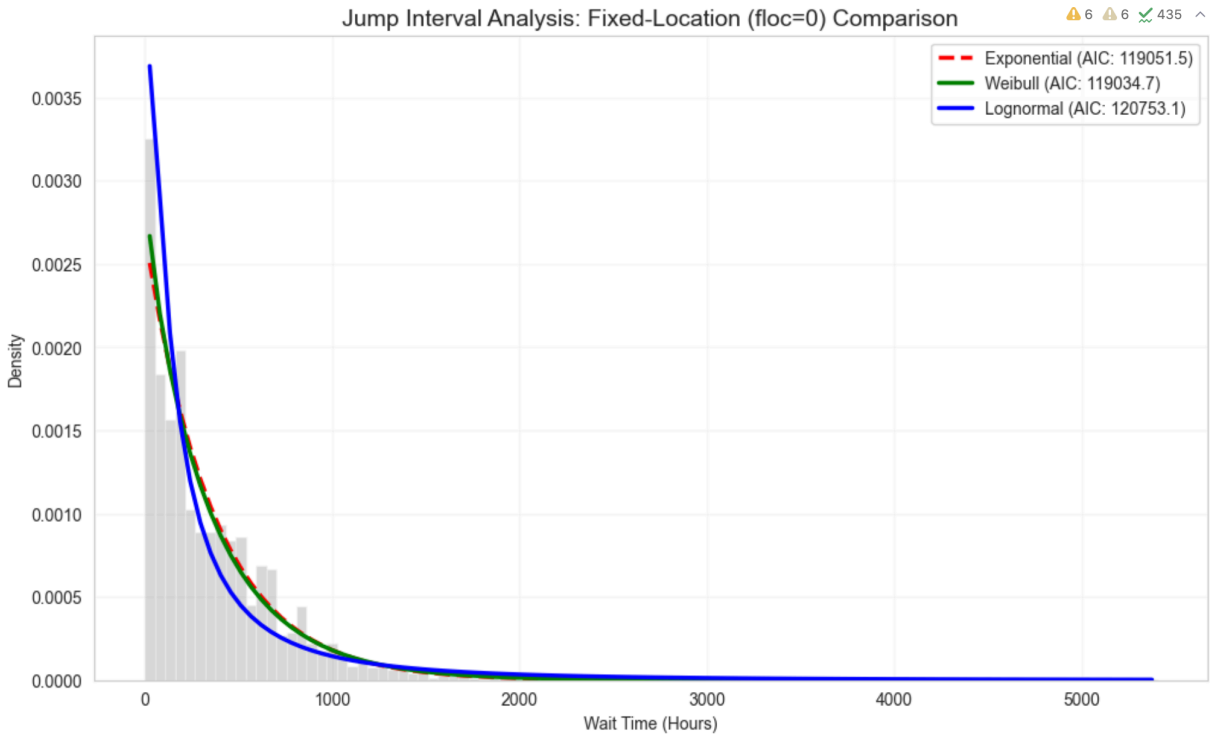
1. Identifying Stock Jumps

I have written the algorithm of identifying jumps in small cap stocks using Lee-Mykland Method. I tested it using the stock price of Harmonic Inc. (HLIT) from 2024 to 2025:



2. How Jump's Time Interval Evolve Overtime

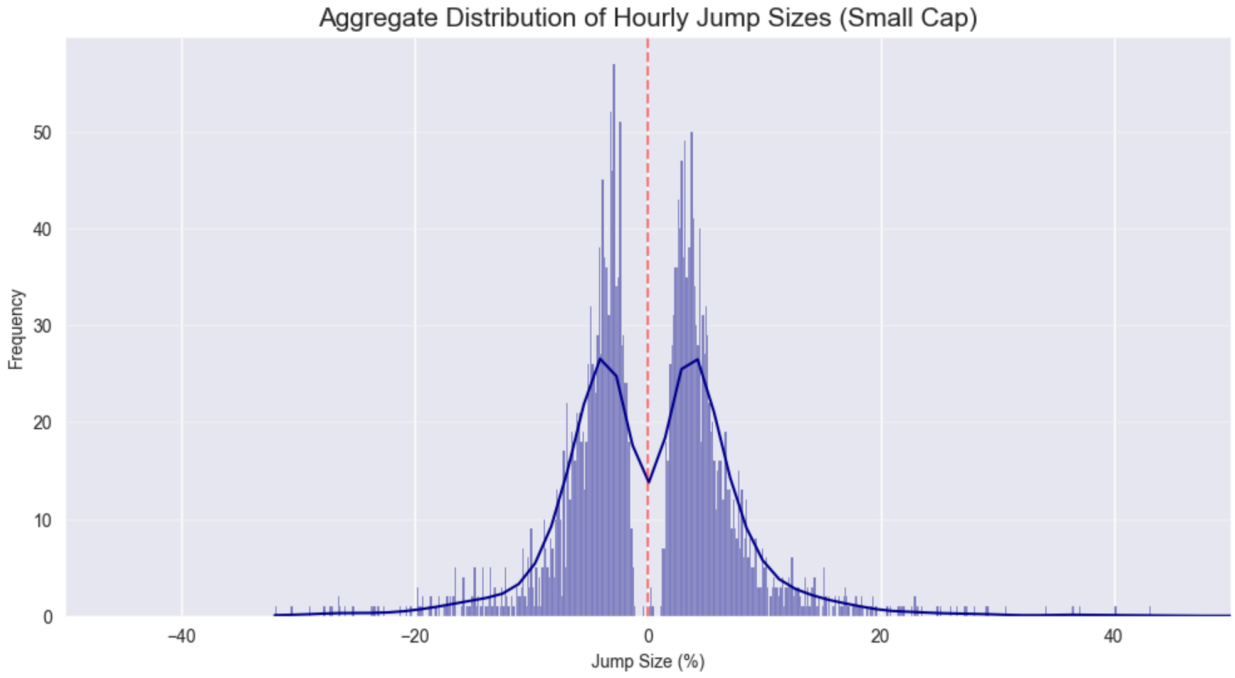
Then, after identifying those jumps, I need to figure out the distribution of the time interval between jumps. I have written the code for doing this, and find out that the distribution of time interval of stock jumps is approximate WeiBull Distribution. This process is different from Merton's theory (who believes that one jump will trigger another jump, which do apply to big cap stocks) and Kou's theory (who believes that the jump diffusion follows a Laplace Distribution). My model's parameters are not fixed, they will evolve overtime, while other models does not, and this process and be difficult to investigate. My results of the time interval:




```
--- Model Selection Criteria (Sorted by AIC) ---
Distribution           AIC           BIC  Log-Likelihood       R2 Shape_Param
     Weibull 119034.705651 119055.887690   -59514.352825 0.951413    0.964248
 Exponential 119051.472839 119065.594198   -59523.736419 0.942600         N/A
   Lognormal 120753.067620 120774.249659   -60373.533810 0.891378     1.35224
```

3. Jump Size Distribution

In this part, I will investigate the jump size distribution of small stocks' jump sizes. After ignoring small volatility jumps (which are not considered as jumps in Lee-Mykland Model), it gives out an dPIN (Double Pareto-Lognormal) Distribution. The results:



4. ML Testing

After doing all of these, I need to test the results, or amend some mistakes or NA values in my jump intensity or probability prediction. Because the jump is determined by factors or signals like Trading Volume, Trading Size, Bid-Ask Spread, Depth Balance, Limit Order Book, Turnover Ratio, etc. While data about these factors may be accessible by yahoo finance and all of these factors are directly proportional to the stock's jumps and may not be that volatile, we can use machines learning to predict their future value. Base on their future value, we can predict the probability or intensity of jumps in the stock's future.

If these values align quite well with previous model, we get our result. Of course, data science and modeling are just tools, theoretical frameworks are vital and I am working on it.

If you need my code of doing previous task, feel free to contact me anytime at LucasFang2@outlook.com or (+86) 178-5799-1820.