International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
OPEN ACCESS
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST2512159

Density Estimation for Financial Fraud Detection: A Multivariate Kernel-Based Approach

Basavaraj Talawar, A. S. Talawar
Department of Statistics, Karnatak University, Dharwad PIN:580 003, Karnataka, India

ARTICLEINFO

Article History:

Accepted : 21 Feb 2025
Published: 23 Feb 2025

Publication Issue :
Volume 12, Issue 1
January-February-2025

Page Number :
676-695

ABSTRACT

The detection of insider trading has become increasingly challenging due to the complexity of modern financial markets. This paper introduces a novel approach using **Volume-Weighted Multivariate Kernel Density Estimation** to identify potential insider trading activities. By integrating volume-weighting with the **Kernel Density Estimation** framework, this method effectively captures abnormal price-volume relationships, providing a sophisticated means to flag irregular trading behaviour. This paper proposes a robust approach using Volume-Weighted Multivariate Kernel Density Estimation to detect abnormal trading patterns linked to insider trading and this approach improves traditional **Kernel Density Estimation** by incorporating trading volume as a weighting factor, allowing it to capture the joint distribution of stock returns and trading volumes. The model's application is tested on data from companies targeted by Hindenburg Research, including Nikola Corporation, Clover Health, and Adani Enterprises. The results show clear pre-event anomalies, including increased trading volumes and price movements preceding the report releases. By utilizing adaptive bandwidth selection, Volume-Weighted Multivariate Kernel Density Estimation balances bias and variance to refine detection of insider trading activities in both dense and sparse regions of the data.

The emergence of reports from activist short-sellers like Hindenburg Research has triggered significant market reactions, often accompanied by allegations of insider trading. This methodology is highly sensitive to abnormal volumes that may signal preemptive insider trading ahead of major market events, such as the release of damaging reports. This research demonstrates that VW-MKDE offers enhanced sensitivity to volume-weighted deviations, enabling the detection of abnormal trading behavior associated with potential insider information. Our findings highlight the tool's capacity to assist regulators and analysts in identifying suspicious market activities in real-time, particularly during periods of heightened market volatility and post-report market responses.

Keywords: Insider trading, volume-weighted Kernel density, Hindenburg report, market manipulation, adaptive bandwidth.

I. INTRODUCTION

Detecting insider trading in modern financial markets presents significant challenges, particularly as the complexity and volume of trading activity increase. Traditional detection techniques, while effective in simpler market conditions, often fail to capture these subtle abnormalities, particularly in high-volume environments where insider trading may be hidden within large-scale legitimate trades. To address these limitations, this paper introduces a Volume-Weighted Multivariate Kernel Density Estimation (VW-MKDE) method. Kernel density estimation (KDE) is a non-parametric technique used to estimate the probability density function (PDF) of a random variable (Acerbi and Tasche, 2002). However, the traditional KDE approach is often less effective when applied to financial data, as it doesn't account for the varying significance of trading volumes, which are critical indicators of market activity. The VW-MKDE improves upon the traditional KDE by incorporating trading volume as a weight, thereby emphasizing trades with larger volumes that may signify insider activity. This weighted approach enhances the model's ability to detect abnormalities in both price and volume data, making it more sensitive to unusual market behaviors (Aggarwal and Wu, 2003). This method is particularly valuable for identifying pre-event trading anomalies, especially surrounding the release of market-moving reports from activist short-sellers, such as Hindenburg Research. These reports frequently lead to significant market reactions and raise the potential for insider trading (Artzner, et al.,1999). By employing the VW-MKDE, we can better capture joint distributions of stock returns and trading volumes, allowing for the detection of abnormal price-volume relationships that would otherwise be overlooked by traditional models (Biswas and Sen, 2018).

A key feature of this approach is the use of **adaptive bandwidth selection**, which balances the trade-off between bias and variance. In KDE, the choice of bandwidth significantly influences the smoothness of the density estimate. An **adaptive bandwidth** allows the model to adjust to areas of high and low data density, using smaller bandwidths in dense regions to capture fine details and larger bandwidths in sparse regions to avoid overfitting. This flexibility is critical in financial data, where market conditions can vary dramatically within short time periods. To demonstrate the efficacy of VW-MKDE, and the model to data from companies such as Nikola Corporation, Clover Health, and Adani Enterprises (Bolancé et al., 2011). These companies were targeted by reports from Hindenburg Research, and their stock prices and trading volumes exhibited significant pre-event anomalies that may indicate insider trading (Brandtner, 2019). Our analysis reveals that VW-MKDE is highly sensitive to these anomalies, detecting increased trading volumes and price deviations in the period leading up to the release of these reports This contributes to the field of market regulation by providing a more nuanced tool for real-time detection of insider trading. The VW-MKDE's ability to handle large, complex datasets with varying trade volumes makes it particularly suited for modern financial markets, where high-frequency and algorithmic trading dominate. By focusing on volume-weighted deviations, this method offers a more sophisticated means of identifying suspicious trading patterns, helping regulators and market analysts maintain market integrity in the face of increasingly complex trading strategies (Haug et.al, 2013). Insider trading, characterized by trades made on material non-public information, can lead to abnormal fluctuations in both stock prices and trading volumes (Ameis, et al., 2023).

II. METHODS AND MATERIAL

2.1 Multivariate Kernel Density Estimation (MKDE)

Multivariate KDE is used to estimate the probability density function of a random variable in multiple dimensions. Consider a d-dimensional random vector $X = (x_1, x_2, \dots, x_d)'$ where $x_1, x_2, \dots, x_d$ are one - dimensional random variables. Drawing a random sample of size "n" in this setting means that we have "n" observations for each of the d random variables $x_1, x_2, \dots, x_d$. Suppose that we collect the i^{th} observation of each of the d random variables in the vector X_i.

$$X_i = (x_1, x_2, \dots, x_d)' \ \forall \ i = 1(1)n$$

From univariate KDE experiences with the one-dimensional case, we might consider adapting the Kernel density function of $X = (x_1, x_2, \dots, x_d)'$, which is a d-dimensional case

$$\widehat{F_h(x)} = \frac{1}{n\,h^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

Where, K is multivariate Kernel function operating and h is a vector of bandwidth, $(h_1, h_2, \dots, h_d)$.

The multivariate Kernel density estimator becomes

$$\widehat{F_h(x)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} h_j^{-1}\ K\left(\frac{x_j - X_{ij}}{h_j}\right)\right\}$$

(Peña, 2012).

2.2 Kernel Function

The Kernel function $K(.)$ is symmetric non-negative function that integers to 1. It controls the shape of the density around each data point. Common choices for the Kernel function include:

panechinkov Kernel: $K(U) = \frac{3}{4}(1 - u^2) \ \ \forall \ |U| \le 1$

2.3 Bandwidth Selection

Consider selecting bandwidths based on the mean integrated squared error (MISE) criteria. The MISE of $\widehat{f_h}$ can be decomposed into a sum of an integrated square bias term and an integrated variance term as below:

$$MSE\left[\widehat{F_h(x)}\right] = \left\{bias_h\left(\widehat{F_h(x)}\right)\right\}^2 + Var\left(\widehat{F_h(x)}\right)$$

2.3.1 Bias of MKDE

$$bias_h\left(\widehat{F_h(x)}\right) = E\left[\widehat{F_h(x)}\right] - \widehat{F_h(x)}$$

$$E\left[\widehat{F_h(x)}\right] = \frac{1}{h^d}\ E\left[K\left(\frac{(x - X_i)}{h}\right)\right]$$

$$Bias_h\left(\widehat{F_h(x)}\right) = \frac{h^2}{2}\int_{R^d} K(U)u^T u\ H\big(f(x)\big)du$$

Therefore, the bias of MKDE is proportional to h^2. Meaning that as the bandwidth h increases, the bias will increase (Brandtner, 2019).

2.3.2 Variance of MKDE

The variance of MKDE is the expected value of the squared difference between the estimate $\widehat{f_h}(x)$ and its expected value.

$$Var\left[\widehat{F_h(x)}\right] = E\left[\left(\widehat{F_h(x)} - E\left[\widehat{F_h(x)}\right]\right)\right]^2$$

$$E\left[\left(\widehat{F_h(x)}\right)^2\right] = \frac{1}{n{h^d}^2}\int_{R^d} K^2(U)f(x - hu) - h^d du$$

$$= E\left[K^2\left(\frac{x - X_i}{h}\right)\right]$$

$$V\left[\widehat{F_h(x)}\right] = \frac{1}{n}\int \{K_H(u - x)\}^2 du - \frac{1}{n}\left\{E\left[\widehat{F_h(x)}\right]\right\}^2$$

$$V\left[\widehat{F_h(x)}\right] = \frac{1}{n}h^d \int_{R^d} K^2(U)f(x)du$$

Thus, Variance of the MKDE is inversely proportional to nh^d. This means that as h increases, the variance decreases and vice -versa.

$$MISE\left[\widehat{F_h(x)}\right] = \int \left\{E\left[\widehat{F_h(x)}\right] - f(x)\right\}^2 dx + \int Var\left(\widehat{F_h(x)}\right)dx$$

Where, if we take equal weights for all data points such that $w(\cdot) = \frac{1}{n}$ and

$$E\left[\widehat{F_h(x)}\right] = \sum E[w(X_i, Z_i)\ K_h(x - X_i)]$$

This bias is $E\left[\widehat{F_h(x)} - f(x)\right]$ can be written as

$$bias_h(x) = \int n\, w(x-ht,z)\; K(t)\, f(x-ht)dt - f(x)$$

By using Taylor's expansion around $f(x)$. The first two terms of the expansion are

$$f(x-ht) = f(x) - ht\, f'(x) + \frac{h^2t^2}{2}\, f''(x)$$

Simplify equation (5) we get

$$bias_h(x) = \left[\int n\, w(x-ht,z)K(t)f(x)dt - \int n\, w(x-ht,z)\, K(t)ht\, f'(x)dt + \int n\, w(x-ht,z)K(t)\frac{h^2t^2}{2}f''(x)dt - f(x)\right]$$

$$bias_h(x) = \left[f(x)\left[n\int w(x-ht,z)K(t)dt - 1\right] - nhf'(x)\int w(x-ht,z)K(t)dt + \frac{nh^2}{2}\, f''(x)\int t^2w(x-ht,z)k(t)dt + O(h^3)\right]$$

If $w(\cdot) \neq \frac{1}{n}, n.w(x-ht,z)f(x) - f(x) \neq 0$ and $\int tK(t)dt = 0$

A rough approximation for handling weighted samples is to expand them into an equivalent unweighted sample of the same size, where censored observations are replaced with imputed values. This allows the use of the standard optimal bandwidth selection method, $h_{\{opt\}}$, to estimate density while treating all data points with equal importance (Breunig, 2006).

2.3.3 Assumption based Approach (Epanchinkov Kernel)

$\therefore K(U) = \frac{3}{4}(1-u^2)\ \forall\ |u| \leq 1$ Is outside the range $|u| > 1, K(U) = 0$ which makes it a compact Kernel meaning it has a limited range.

Bandwidth: The bandwidth h controls the smoothing for the Epanechinkov Kernel it applies:

- ❖ If small h reduces bias the increases variances.
- ❖ If large h reduces variances, then increases bias.

Bias: The bias of MKDE with a Kernel $K(U)$ is typically proportional to h^2 i.e.

$$Bias \propto h^2.f''(x)$$

Therefore, this bias is relatively smaller compared to the other Kernels (Gaussian or Uniform) because it is optimized MISE.

Bias-variance trade-off: To minimize bias want to select an optimal band h which balances the trade-off between bias and variance. As a rough approach, the rule of thumb bandwidth can be used

$$h = 2.34\ \hat{\sigma}\, n^{-\frac{1}{5}}$$

Where $\hat{\sigma}$ is the standard deviation and n is the total number of observations.

2.3.4 Optimal Bandwidth

To minimize the MISE, we can derive the optimal bandwidth h_{opt} by balancing the bias and variance terms. Taking the derivative of the MISE w.r.t h and setting it to zero gives.

Therefore, we consider H *and* $\sum_d$ to be diagonal matrices $H = diag(h_1, h_2, \dots, h_d$ and $\sum = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2)$ these leads

$$\tilde{h} = \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}}\ \frac{n^{-1}}{d+4}\ \widehat{\sigma_d}$$

$$h_{opt} \propto n^{-\frac{1}{d+4}}\ or\ h_{opt} \propto \left(\frac{1}{(n)^{\frac{1}{d+4}}}\right)$$

It shows that the optimal bandwidth h needs to be smaller, but the risk of overfitting increases as dimensionality grows, leading to the curse of dimensionality.

2.3.5 Plug-in Estimator

A plug in method can be used to automatically select the bandwidth by estimating $\int f''^2$ in

$$h_{opt} = K_2^{-\frac{2}{5}}\left\{\int K(t)^2dt\right\}\left\{\int f''(x)^2dx\right\}^{-\frac{1}{5}} n^{-\frac{1}{5}}$$

With another Kernel based estimate as in

$$\widehat{f_h}(x) = \sum_{i=1}^{n} w(x_i,z_i)[K_h(x-X_i) + K_h(x+X_i) * I_{0\leq x<4h}]$$

However, for weighted samples or incomplete data such as the right censored data. We won't be able to get a close form of the bandwidth h by minimizing the first two terms of the usual asymptotic expansion of its MISE.

$$AIMSE\big(E(h)\big) = h^{-1}K(U) + \frac{h^4\sigma_K^4 K(f'')}{4}$$

so, the direct plug in method won't work here so we will move least squares cross validation method (Rizvi et al., 2017).

2.4 Least squares cross validation method

To select the bandwidth completely automatically, we apply the least squares cross validation (LSCV). LSCV minimize the integrated squared error (ISE) between the true density $f(x)$ and the MKDE $\widehat{f_h}(x)$. Since the true density is unknown, LSCV minimizes an estimate of the MISE.

The LSCV objective function is given by

$$LSCV(h) = \int \widehat{f_h}(x)dx + \int f^2(x)dx - 2\int \widehat{f_h}(x)f(x)dx$$

The Epanechinkov Kernel can be approximated of first term in above equation becomes

$$\int \widehat{f_h^2}(x)\,dx = \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right)^2$$

The final LSCV function to minimize can be written as

$$LSCV(h) = \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right) - \frac{2}{n}\sum_{i=1}^{n}\frac{1}{(n-1)h}\sum_{\substack{j=1\\ j\neq i}}^{n} K\left(\frac{X_i - X_j}{h}\right)$$

For the Epanechinkov Kernel, this integral has known as solution and its value depends on the bandwidth h and distance between X_i and X_j.

$$\widehat{f_{h_{1-\iota}}}(x) = \frac{1}{(n-1)h}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{3}{4}\left(1-\left(\frac{(X_i - X_j)}{h}\right)^2\right)$$

$$\forall\left|X_i - X_j\right| \le h$$

Here, the weight may not be necessary be sum up to one. So, we simply leave the denominator as it.

2.5 Adaptive weighted Multivariate Kernel Density Estimator (AWMKDE)

The **AWMKDE** is an advanced extension of the traditional kernel density estimation (KDE) method, designed to handle complex data distributions by adapting bandwidths locally and assigning varying importance to data points. Let's go through a detailed explanation with mathematical formulations (Wächter, et al., 2010).

2.5.1 Adaptive Bandwidth

To address this, adaptive bandwidth KDE allows the bandwidth to vary for each data point X_i. One common approach is to define h_i using a pilot density estimate $\tilde{f}(X_i)$, which is an initial density estimate obtained using a fixed bandwidth. Then:

$$h_i = h\left(\frac{\tilde{f}(X_i)}{g}\right)^{-\alpha}$$

Where, g is the geometric mean of $\tilde{f}(X_i)$, α is a sensitivity parameter and h is global bandwidth.

2.5.2 Weights (W_i)

These weights can be determined based on a variety of factors. A common approach is to assign higher weights to points that are closer to x importance of X_i in the context of the application.

The weight of MKDE is given by:

$$W_i = \frac{\left(\frac{K(x-X_i)}{h_i}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_j}\right)}$$

This normalizes the weights, ensuring that they sum to 1.

The Adaptive Weighted Multivariate Kernel Density Estimation (AWMKDE) enhances density estimation by using adaptive bandwidths and weighted observations, allowing for precise modelling of complex, non-uniform data distributions in both dense and sparse regions (Wang and Wang, 2019).

2.5.3 Volume-Weighted KDE in Insider Trading Detection

The volume of trades is a critical factor in detecting insider trading. Insiders tend to trade in ways that reflect their private information, often leading to abnormal volumes or price patterns. Volume-Weighted KDE gives higher importance (weight) to trades that involve larger volumes, as these could signify trades driven by privileged information (Haug, et al., 2013).

In this approach:

- Higher trading volumes might be weighted more heavily, reflecting their potential importance in insider trading detection.
- Smaller trading volumes may have less weight in the KDE, as they are less likely to represent significant, information-driven trades.

The Volume-Weighted KDE formula can be expressed as:

$$\widehat{f_h}(x) = \sum_{i=1}^{n} w_i \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right)$$

Where:

- w_i represents the volume weights, proportional to the trading volume at X_i,
- X_i could represent trade-specific variables like time, price, volume, and trade direction (buy/sell),
- h_i is the adaptive bandwidth, which can depend on the density of trades in that region.

2.5.4 Adaptive Bandwidth for Volume-Weighted KDE

In detecting insider trading, the adaptive bandwidth h_i allows for finer detection in dense areas (Regions with normal trading activity) and smoother detection in sparse regions (Regions where abnormal activity might be happening).

- In normal market conditions: (High trading activity), Smaller bandwidths h_i allow detecting subtle patterns, such as unusual price movements relative to the volume.
- In abnormal conditions: (Sparse trading activity), Larger bandwidths smooth out the noise, making it easier to detect large deviations or clusters of suspicious activity.

Volume-weighted KDE Formula for insider trading be formulated as:

$$\widehat{f_h}(x) = \sum_{i=1}^{n} w_i \left(\frac{1}{h_i^4}\right) K\left(\frac{V - V_i, P - P_i, T - T_i, D - D_i}{h_i}\right)$$

Where, V_i, P_i, T_i *and* D_i represent the trade volume, price change, time, and trade direction at observation i, w_i represents the volume weight, calculated based on the trading volume V_i such that larger volumes get higher weights, h_i is the adaptive bandwidth, varying locally based on the density of data around X_i (Efromovich, 2009).

The AWMKDE approach, when combined with volume weighting, is highly suited for insider trading detection because it provides more weight to high-volume trades, which are more likely to be driven by insiders with private information. By using volume-weighted KDE with adaptive bandwidth selection, regulators or market analysts can detect abnormal trading patterns that warrant further investigation (Frees, 2004).

III.RESULTS AND DISCUSSION

3.1 Stock Price and Trading Volume Data

Data obtained from Yahoo! Finance daily stock prices and trading volumes for companies targeted by Hindenburg Research reports.

Table 1: Nikola Corporation price and volume data

Date	Open	High	Low	Close	Volume
09/21/2020	10.07	10.333	10.07	10.2	1441900
09/22/2020	10.25	10.5	10.192	10.33	212300
09/23/2020	11.36	11.36	10.25	10.4	410300
……	……	…..	……	…..	……
08/16/2023	3.16	3.33	3.03	3.1	7746700
08/17/2023	3.1	3.179	2.79	2.82	12844800
08/18/2023	2.81	2.83	2.54	2.63	18536700

Table 2: Clover Health prices and volumes dataset

Date	Open	High	Low	Close	Volume
01/1/2018	90.34	92.55	89.90	90.48	7745247
01/2/2018	90.50	91.21	88.35	89.30	4977129
01/3/2018	89.30	98.21	89.30	96.01	39032205
……	……	…..	……	…..	……
01/25/2023	3422	3428	3315	3388.95	3686439
01/27/2023	3335	3346.5	2712	2761.45	14764368
01/30/2023	2850	3037.55	2665	2892.85	20947906

Table 3: Adani Enterprises LTD prices and volumes dataset.

Date	Open	High	Low	Close	Volume
9/21/2020	309.30	309.60	309.30	309.60	2010
9/22/2020	309	309	309	309	500
9/23/2020	309.75	310.50	309.60	309.60	2290
……	……	…..	……	…..	……
12/6/2023	7.20	7.21	6.47	6.51	3789600
12/7/2023	6.60	6.73	6.51	6.57	1415800
12/8/2023	6.57	6.71	6.43	6.63	2022900

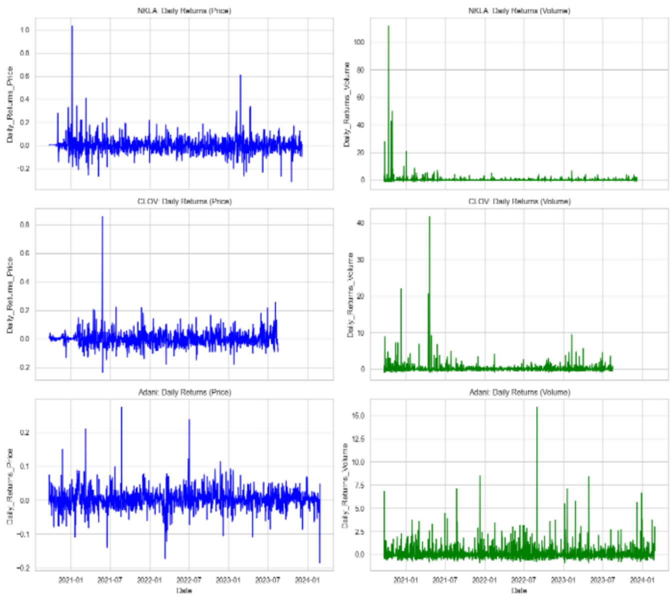


Figure 1: Pattern observation of all datasets

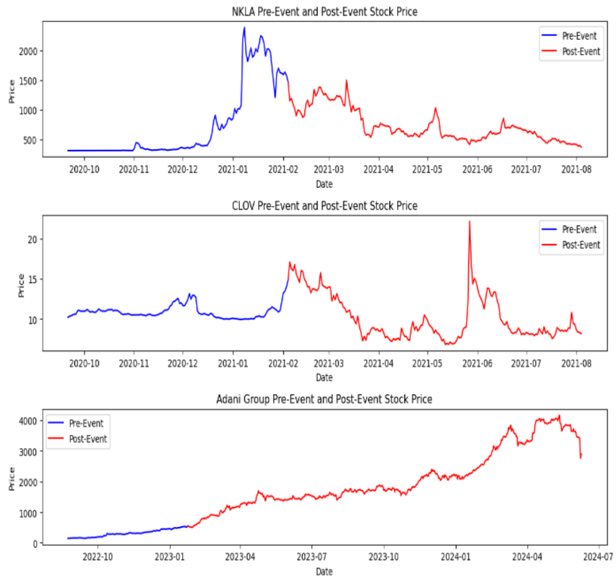
3.2 Event Date Information

The Hindenburg Research reports were released dates are:

- Nikola Corporation – 10 September 2020
- Clover Health – 04 February 2021
- Adani Enterprises Ltd – 24 January 2023

3.3 Time Window

A time window of several months before and after the release of report to analyze pattern.



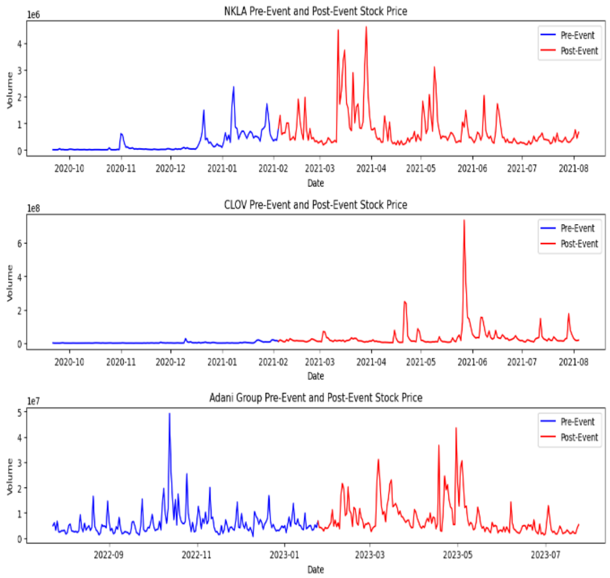


Figure 2: A time window of pattern variation of daily returns and volumes

The stock price trends of NKLA, CLOV, and Adani Enterprises before the Hindenburg reports suggest speculative or steady growth, with NKLA showing the most pronounced pre-event rise. Post-event reactions vary: NKLA and CLOV experienced sharp declines, while Adani Enterprises exhibited continued growth with increased volatility rather than an immediate drop. Nikola Corporation (NKLA) and Clover Health (CLOV) show sharp negative reactions in price with increased volume right after the Hindenburg report release, suggesting immediate negative market sentiment. Adani Group (Adani Enterprises Ltd) shows more gradual price and volume changes, with less dramatic immediate market impact but with increased volatility and trading activity post-report. This behavior is typical in markets where new negative information, like allegations of fraud, drastically changes investor sentiment, leading to heavy trading volumes and price volatility.

3.4 Daily Returns, Normalized Trading Volumes and Multivariate Dataset

Compute daily returns for the stock as:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Where: R_t is the return at time t, P_t is the price at time t, and P_{t-1} is the price at time t−1. The three datasets (Nikola, Clover, Adani), we would calculate the daily returns for each individually.

To normalize stock prices (or any data) means adjusting the values so they are on a consistent scale, often to compare trends between different stocks. There are various ways to normalize data, but one of the simplest and most common methods is Min-Max normalization:

$$V_t = \frac{\text{Volume}_t}{\text{Max}(\text{Volume}_t)}$$

Normalize trading volumes: Min-Max normalization scales your data between 0 and 1, helping to compare trends across stocks. Initial value normalization shows how the price changes relative to the starting price (usually helpful for visualizing percentage gains or losses over time).

$$Normalized\ value = \frac{P_t - P_{\min}}{P_{\max} - P_{\min}}$$

Where, P_t is the price at time t, $P_{\min}$ is the minimum price in the dataset, and $P_{\max}$ is the maximum price in the dataset.

Daily Returns: This column represents the percentage change in the stock's closing price from the previous day. Positive values indicate a gain, and negative values represent a loss. For instance, on 2020-08-05, the stock increased by approximately 6.67%, while on 2020-08-06, it dropped by about 3.13%.

Normalized Volume: This column reflects the trading volume scaled between 0 and 1 using min-max normalization. A value of 1.00 indicates the highest volume (on 2020-08-10), and 0.25 represents a lower volume (on 2020-08-05).

Table 4: The daily returns and normalized values of Nikola corporation

Date	Daily Returns	Normalized Values
2020-08-05	0.066667	0.25
2020-08-06	-0.031250	0.50
2020-08-07	0.096774	0.75
2020-08-10	-0.029412	1.00

Table 5: The daily returns and normalized values of Clover Health

Date	Daily Returns	Normalized Values
2020-08-05	0.166667	0.25
2020-08-06	-0.071429	0.50
2020-08-07	0.153846	0.75
2020-08-10	0.066667	1.00

Table 6: The daily return and normalized value of Adani Ent Ltd.

Date	Daily Returns	Normalized Values
2021-02-02	0.222222	0.00
2021-02-03	0.090909	0.17
2021-02-04	0.166667	0.67
2021-02-05	0.071429	1.00

These multivariate datasets allow us to assess how price returns and trading volumes are interrelated during critical periods for the stock.

3.5 MKDE Implementation

3.5.1 Kernel Choice: The Epanechnikov KDE unveils a striking concentration of market activity near the origin, with Nikola Corporation leading in density, as its finely tuned bandwidth balances precision and efficiency, transforming raw data into a powerful lens for decoding market behavior and risk dynamics.

3.5.2 Bandwidth Selection: Silverman's Rule of Thumb fine-tunes MKDE bandwidths for NKLA, CLOV, and Adani Enterprises, where the Epanechnikov Kernel excels in anomaly detection, the Uniform Kernel exaggerates outliers, and Volume-Weighted MKDE unveils CLOV's amplified sensitivity to volume shifts, making it an indispensable lens for decoding market turbulence.

Table 7: Volume Weighted MKDE values and Optimal bandwidth.

		NKLA	MKDE Values at the point			CLOV
Optimal Bandwidth	Gaussian	0.16	0	0	0	0.10
	Epanec-inkov	0.10	0.84	0.57	0.02	0.10
	Uniform	0.10	5.45	5.45	5.70	0.10
Volume-Weighted			0.17	0.07	0.13	

MKDE Values At the point			ADANI ENT.NS	KDE Values at the point		
0	0	0	0.10	0	0	0
0.78	0.57	0.06	0.10	0.75	0.49	0.09
0	0	0	0.10	0.07	0.07	0.07
0.73	0.55	0.77		0.46	0.29	0.43

Overall, the analysis underscores the importance of tailored bandwidth selection and kernel choice to capture key market dynamics effectively.

3.6 Volume Weighted Multivariate KDEs

Leveraging NumPy and SciPy, the Volume-Weighted Multivariate KDE refines probability density estimation by integrating trading volume as a dynamic weight, unveiling intricate return-volume dependencies and exposing deep market dynamics through striking contour and 3D visualizations

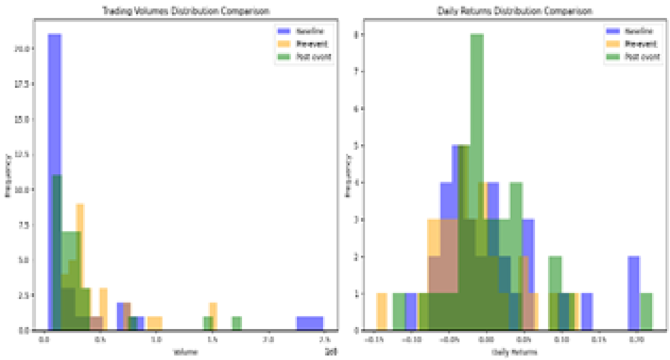


Figure 3: The joint distribution of stock returns and trading volumes of NKLA.

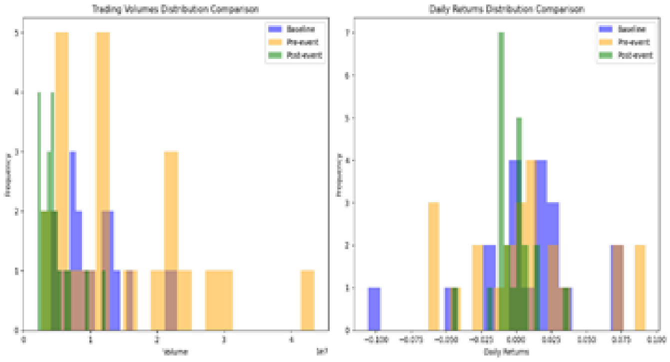


Figure 4: The joint distribution of stock returns and trading volumes of CLOV

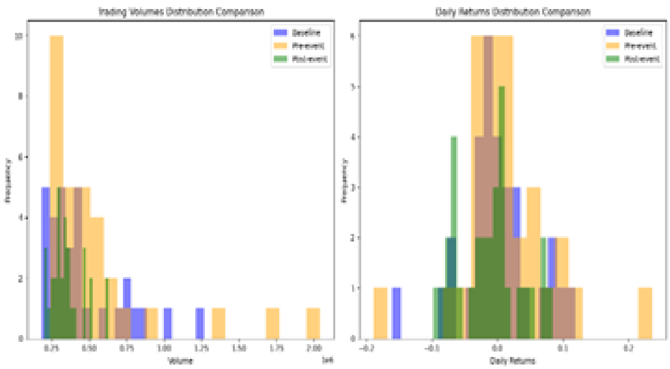


Figure 5: The joint distribution of stock returns and trading volumes of ADANIENT.NS

The Hindenburg reports acted as seismic shocks to the market, transforming a pre-event calm into a post-event storm of volatility, with trading volumes and returns unraveling in a chaotic yet revealing display of investor panic and market recalibration.

3.7 Abnormal Trading Detection

3.7.1 Threshold for Abnormal Behavior: To detect abnormal days based on Multivariate Kernel Density Estimation (MKDE), We establishing an abnormal threshold to identify days, where the estimated density of returns and trading volumes exceeds a specific threshold. The goal is to flag days where the joint distribution of returns and volumes deviates significantly from the expected norm.

3.7.2 Detect Insider Trading:

Evidence suggests possible insider trading, with unusual volume spikes and price movements before the reports' release. Elevated pre-event trading activity hints at market speculation, while post-release periods saw sharp price declines and high volumes as investors reacted. Extreme volatility followed, with sustained trading activity as markets adjusted to new information.

Table 8: Potential Insider Trading Cases of NKLA

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
1	2020-09-22	-0.001938	500	True	35	2020-10-26	-0.003813	3303	True
2	2020-09-23	0.001942	2290	True	36	2020-10-27	-0.004785	63	True
3	2020-09-24	0.000969	1430	True	37	2020-10-28	-0.000961	37	True
5	2020-09-26	0.000774	5623	True	38	2020-10-29	0.000674	320	True
6	2020-09-27	-0.001932	12333	True	40	2020-10-31	-0.001929	6670	True
7	2020-09-28	0.000968	853	True	42	2020-11-02	0.278261	556077	True

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
8	2020-09-29	0.000967	333	True	45	2020-11-05	-0.144105	104993	True
9	2020-09-30	0.000000	0	True	58	2020-11-18	0.009132	12783	True
12	2020-10-03	0.000967	11890	True	59	2020-11-19	-0.015385	9127	True
13	2020-10-04	0.000000	1113	True	60	2020-11-20	0.008732	7607	True
14	2020-10-05	0.000000	6000	True	63	2020-11-23	0.019830	7500	True
15	2020-10-06	-0.000966	11667	True	64	2020-11-24	0.000000	5997	True
16	2020-10-07	0.000000	3333	True	65	2020-11-25	0.018519	14953	True
17	2020-10-08	0.000484	3333	True	69	2020-11-29	0.056655	38667	True
18	2020-10-09	0.000483	3933	True	72	2020-12-02	-0.008475	8807	True
19	2020-10-10	-0.000193	8550	True	73	2020-12-03	0.003419	10843	True
20	2020-10-11	-0.000773	87	True	75	2020-12-05	-0.038811	19390	True
21	2020-10-12	0.000967	87	True	79	2020-12-09	0.129921	89093	True
22	2020-10-13	0.000966	9820	True	87	2020-12-17	0.126707	55543	True
23	2020-10-14	0.000000	213	True	89	2020-12-19	0.234403	368077	True
24	2020-10-15	0.000000	0	True	90	2020-12-20	0.329735	659487	True
25	2020-10-16	0.001931	4760	True	92	2020-12-22	-0.148845	381547	True
26	2020-10-17	0.001927	5170	True	95	2020-12-25	0.153459	287240	True
27	2020-10-18	-0.001923	2570	True	103	2021-01-02	0.184321	296167	True
28	2020-10-19	0.000000	4287	True	108	2021-01-07	1.036975	1667623	True
29	2020-	-0.000963	253	True	129	2021-01-	0.343028	1725087	True

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
	10-20					28			
30	2020-10-21	-0.000964	690	True	137	2021-02-05	-0.212735	1287047	True
31	2020-10-22	0.000965	1690	True	147	2021-02-15	0.216333	866233	True
32	2020-10-23	-0.002893	167	True	183	2021-03-23	-0.258155	1600113	True
33	2020-10-24	0.003868	1683	True	---	------------	-------------	----------	------

Table 9: Potential Insider Trading Cases of CLOV.

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
1	2020-09-22	0.0127	212300	True	302	2021-07-20	0.098522	37674700	True
4	2020-09-25	0.0095	130400	True	429	2021-11-24	0.218906	47203700	True
8	2020-09-29	-0.0117	365500	True	436	2021-12-01	0.181818	48239800	True
9	2020-09-30	0.0036	238800	True	437	2021-12-02	0.150000	43854600	True
10	2020-10-01	-0.0045	271800	True	441	2021-12-06	-0.139159	16324700	True
11	2020-10-02	0.0082	193100	True	484	2022-01-18	0.139738	12073000	True
13	2020-10-04	-0.0100	348300	True	543	2022-03-18	-0.179412	15744000	True
14	2020-10-05	-0.0100	275400	True	579	2022-04-23	-0.121951	13181100	True
15	2020-10-06	0.0082	365800	True	609	2022-05-23	0.138211	12566600	True
16	2020-10-07	-0.0091	286500	True	682	2022-08-04	0.147826	20513300	True
17	2020-10-08	- 0.009141	285700	True	683	2022-08-05	-0.147727	11443600	True
18	2020-10-11	-0.001845	172500	True	749	2022-10-10	0.165158	10266100	True
23	2020-10-14	0.015712	344900	True	762	2022-10-23	0.005669	57743400	True

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
24	2020-10-15	-0.003640	364100	True	796	2022-11-26	0.136364	9900200	True
36	2020-10-27	0.009132	220900	True	808	2022-12-08	0.113821	4752700	True
37	2020-10-28	-0.001887	279400	True	841	2023-01-10	-0.109709	8162100	True
38	2020-10-29	0.006616	319300	True	854	2023-01-23	0.108672	5442800	True
44	2020-11-04	- 0.004695	235900	True	855	2023-01-24	0.079208	3256400	True
48	2020-11-08	-0.008588	226600	True	857	2023-01-26	-0.181416	43103700	True
49	2020-11-09	0.010587	180000	True	862	2023-01-31	0.123471	6409500	True
54	2020-11-14	0.130660	129700	True	982	2023-05-31	0.177027	22113700	True
56	2020-11-16	0.022883	285400	True	1012	2023-06-30	0.213592	24606700	True
271	2021-06-19	0.044229	4981200	True	1049	2023-08-06	0.255000	26263900	True

Table 10: Potential Insider Trading Cases of ADANIENT

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
43	2020-09-22	-0.073171	14797586	True	781	2020-10-26	0.001942	10085652	True
45	2020-09-23	-0.018873	35549750	True	785	2020-10-27	0.000969	535652	True
48	2020-09-24	-0.015408	27822361	True	769	2020-10-28	0.343028	529556	True
107	2020-09-26	0.060465	17145197	True	824	2020-10-29	-0.212735	425422	True
121	2020-09-27	-0.108788	18195059	True	865	2020-10-31	-0.003813	45282656	True
144	2020-09-28	0.153459	11825552	True	882	2020-11-02	-0.004785	524615	True
145	2020-09-29	0.184321	10865044	True	895	2020-11-05	-0.000961	9785462	True
146	2020-	1.036975	7308821	True	897	2020-	0.000674	124686261	True

Days	Date	Daily Returns	Volume	Abnormal Behavior	Days	Date	Daily Returns	Volume	Abnormal Behavior
	09-30					11-18			
158	2020-10-03	0.343028	8147287	True	900	2020-11-19	-0.001929	1254683	True
192	2020-10-04	-0.212735	11122413	True	925	2020-11-20	0.008732	2564276	True
193	2020-10-05	-0.003813	13340428	True	935	2020-11-23	0.019830	2564	True
196	2020-10-06	-0.004785	12691224	True	961	2020-11-24	0.000000	582746464	True
201	2020-10-07	-0.000961	9129656	True	996	2020-11-25	0.343028	12546866	True
208	2020-10-08	0.000674	10018309	True	1011	2020-11-29	-0.212735	568479	True
265	2020-10-09	-0.001929	8936687	True	1025	2020-12-02	-0.003813	564923	True
298	2020-10-10	0.278261	11610335	True	1026	2020-12-03	-0.004785	13594678	True
299	2020-10-11	-0.144105	7621480	True	1024	2020-12-05	-0.000961	46597132	True
356	2020-10-12	0.018519	16437924	True	1085	2020-12-09	0.000674	26594731	True
458	2020-10-13	0.056655	9189024	True	1088	2020-12-17	-0.001929	26594381	True
561	2020-10-14	-0.001938	5353014	True	1190	2020-12-19	0.000968	16497621	True
598	2020-10-15	0.000153	10994042	True	1191	2020-12-20	0.000967	1643285	True
661	2020-10-16	0.000000	7114891	True	1192	2020-12-22	0.000000	12245148	True
662	2020-10-17	-0.000966	6606714	True	1193	2020-12-25	0.068186	12565925	True
663	2020-10-18	0.000000	7142031	True	1228	2021-01-02	-0.063225	26468565	True
665	2020-10-19	0.000484	14797586	True	1230	2021-01-07	-0.058510	4393335	True
701	2020-10-20	0.000483	39901321	True	1242	2021-01-28	-0.054055	3585883	True
702	2020-10-21	-0.000193	35549750	True	1254	2021-02-05	-0.185161	3036157	True

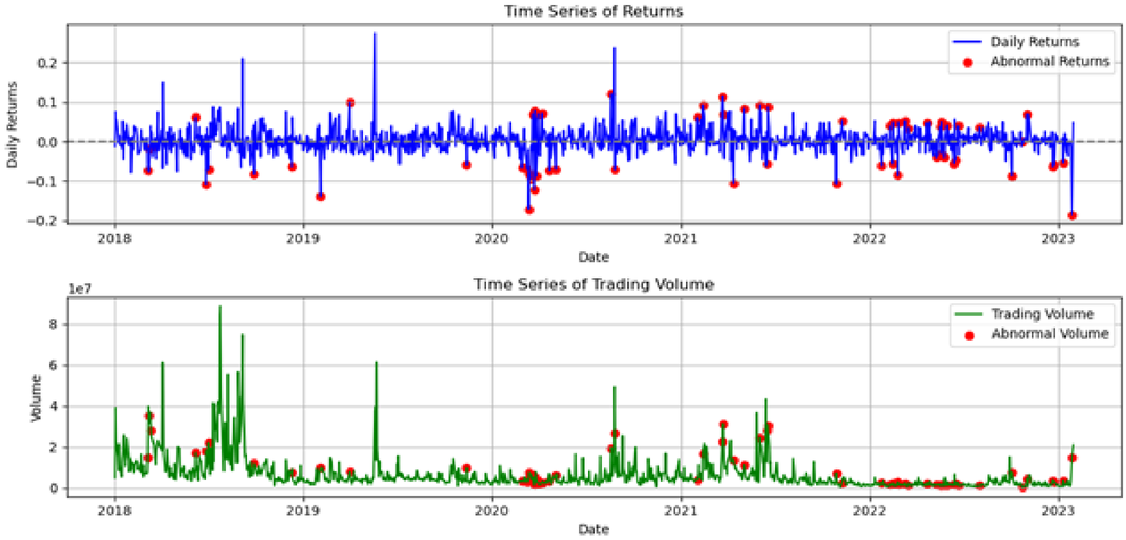


Figure 6: Abnormal behaviours of NKLA

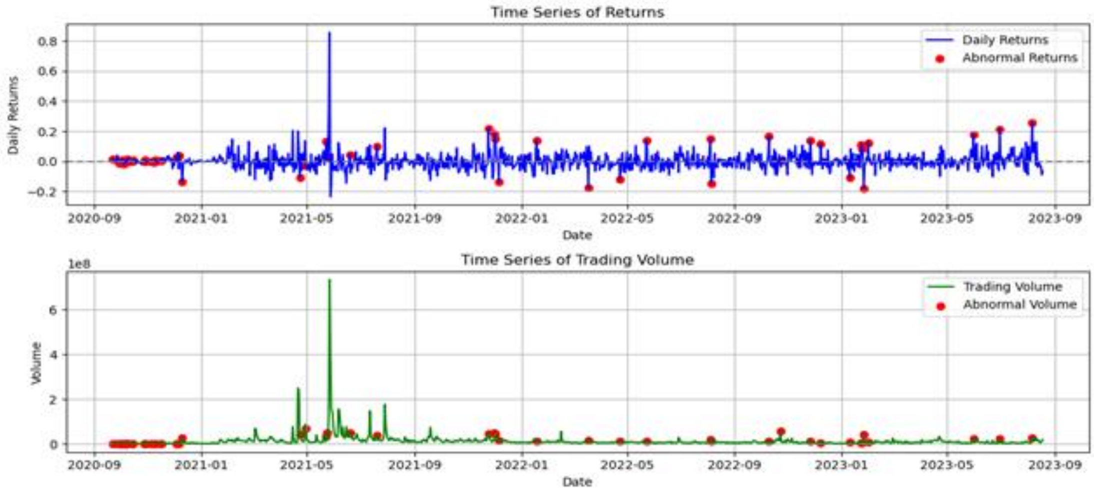


Figure 7: Abnormal behaviours of CLOV

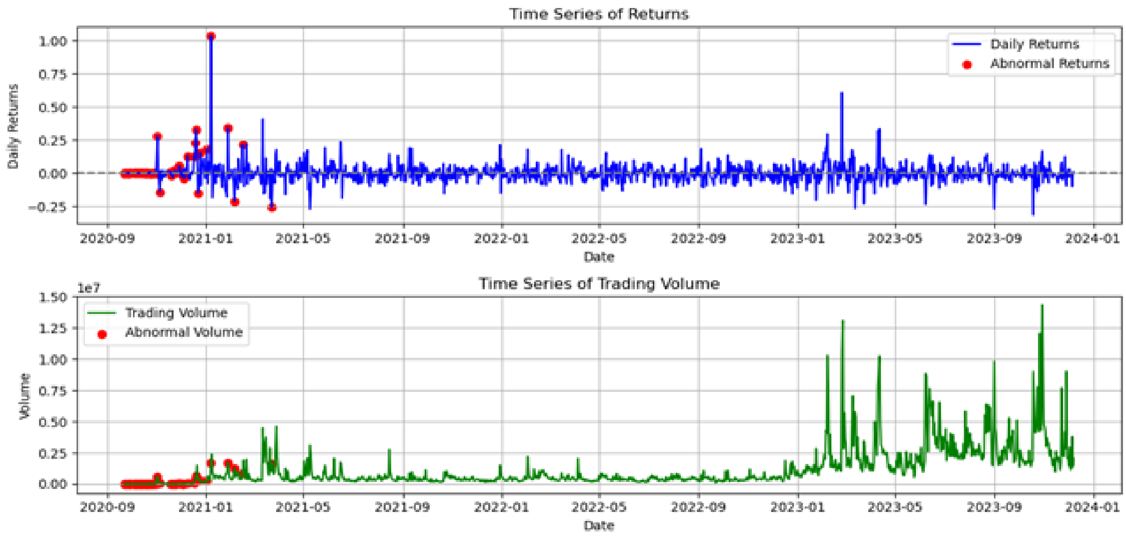


Figure 8: Abnormal behaviors of ADANIENT.NS

Nikola Corporation (NKLA): The surge in abnormal returns and trading volumes around late 2020 suggests a storm of speculative and potentially insider-driven activity, culminating in a market upheaval following the Hindenburg report.

Clover Health (CLOV): The sharp spikes in returns and trading volume around early 2021 paint a picture of a market gripped by speculation, reacting in waves to the Hindenburg report with bursts of volatility.

Adani Enterprises (ADANIENT.NS): The relentless surge in trading activity post-January 2023 underscores a market in turmoil, where investor panic and media scrutiny fueled prolonged volatility and uncertainty.

Abnormal trading behaviors, both in returns and volumes, were concentrated around the release of the Hindenburg report in late 2020. The abnormal spikes in both metrics suggest that the report triggered a significant market reaction, with trading volumes and returns.

3.8 Statistical Testing

3.8.1 Kolmogorov-Smirnov Test:

Table 11: Kolmogorov-Smirnov Test Results for Trading of daily returns and volumes of all three companies.

Companies	Events	Daily Returns		Volumes	
		Statistic	P-Value	Statistic	P-value
NKLA	Pre-event	0.19354	0.61505	0.16129	0.82345
	Post-event	0.16129	0.82345	0.22580	0.82345
CLOV	Pre-event	0.22580	0.41354	0.67741	0.0000462
	Post-event	0.22580	0.41354	0.48387	0.47117
ADANIENT.NS	Pre-event	0.22190	0.58770	0.24256	0.47746
	Post-event	0.39712	0.05622	0.69856	0.0000303

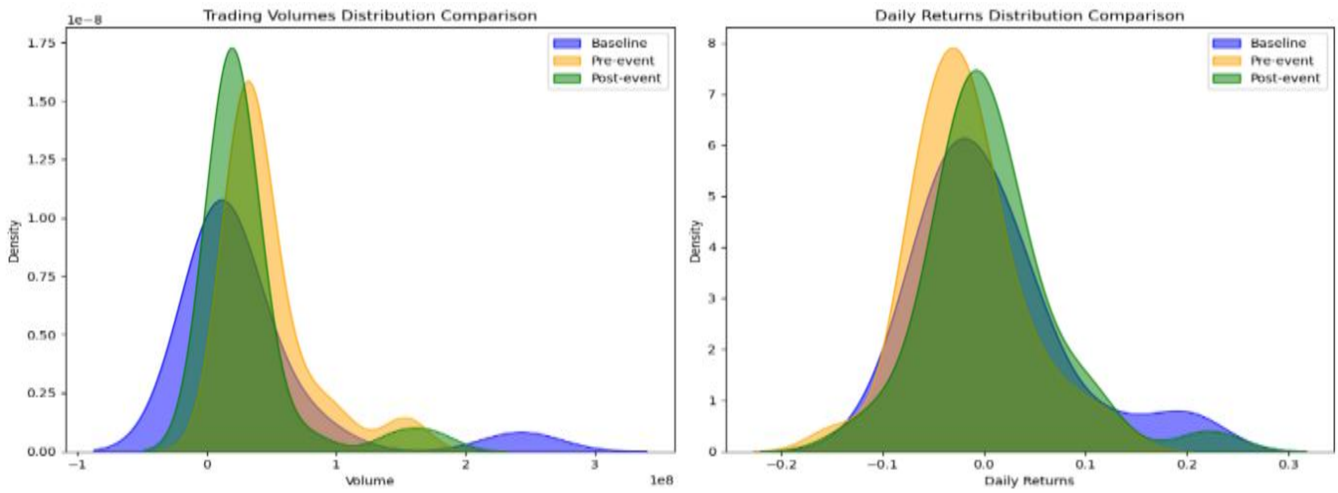


Figure 9: Trading volumes and returns distribution comparison of NKLA

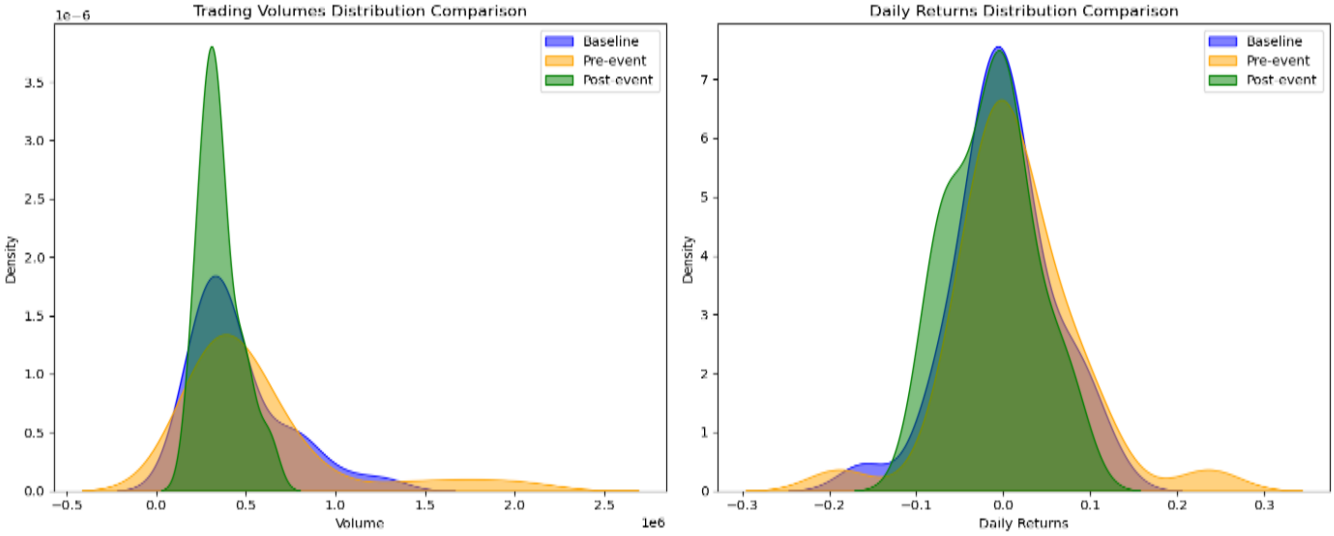


Figure 10: Trading volumes and returns distribution comparison of CLOV

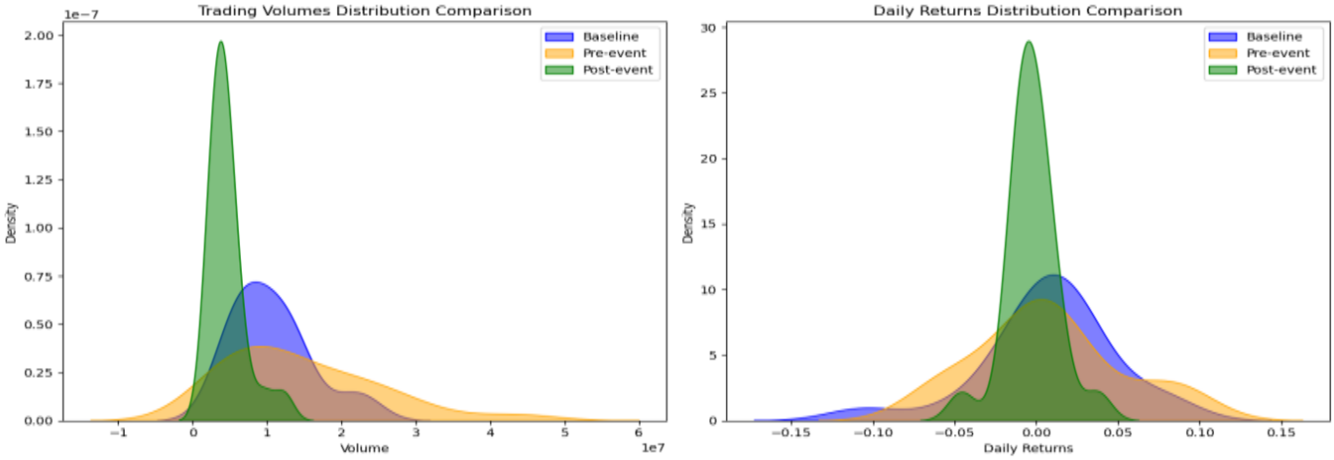


Figure 11: Trading volumes and returns distribution comparison of ADANIENT.NS

NKLA shows minimal post-event impact, CLOV experiences a trading volume surge without major return changes, while ADANIENT.NS undergoes significant shifts in both trading volume and returns, reflecting a stronger market reaction.

Z-Score Calculation

Table 12: Z-Scores for trading volumes with significant deviations.

Events / companies	Pre-Event			Post-Event			Base Line		
	Dates	Volumes	Z-Scores	Dates	Volumes	Z-Scores	Dates	Volumes	Z-Scores
NKLA	08-06-2021	2042073	3.46	13-07-2021	630750	2.46	09-04-2021	1268850	3.06
	16-06-2021	1729870	2.73	22-07-2021	624390	2.40	12-04-2021	1033457	2.17
CLOV	06-06-	154157500	3.03	12-	146920100	3.09	21-	249097800	3.60

Events / companies	Pre-Event Dates	Pre-Event Volumes	Pre-Event Z-Scores	Post-Event Dates	Post-Event Volumes	Post-Event Z-Scores	Base Line Dates	Base Line Volumes	Base Line Z-Scores
	2021			07-2021			04-2021		
	07-06-2021	152247800	2.98	29-07-2021	175897200	2.98	22-04-2021	236758200	3.40
ADANIENT.NS	14-06-2021	43530006	2.81	07-05-2021	12239845	3.17	06-04-2021	21496457	2.05
	------------	-----------	-----	06-07-2021	9485434	2.03	07-04-2021	23007694	2.34

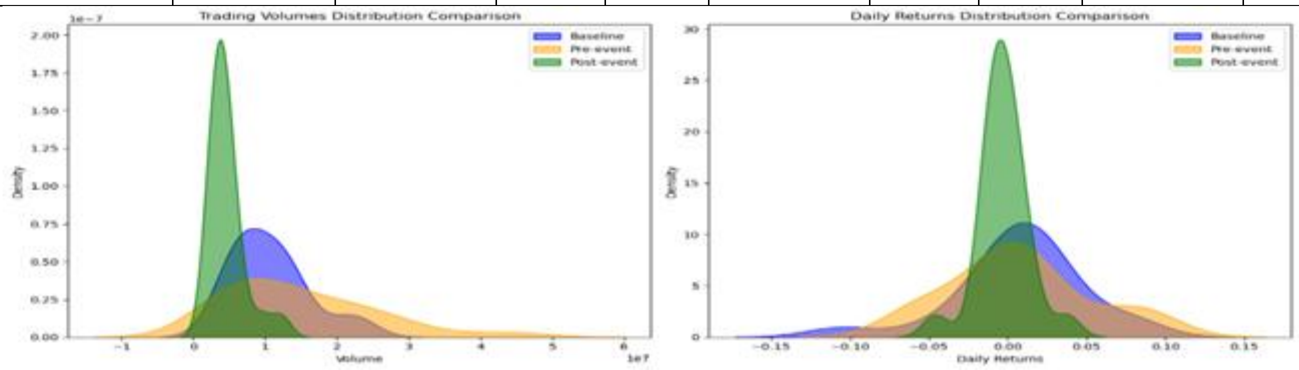


Figure 12: Significant Deviations for trading volumes of NKLA

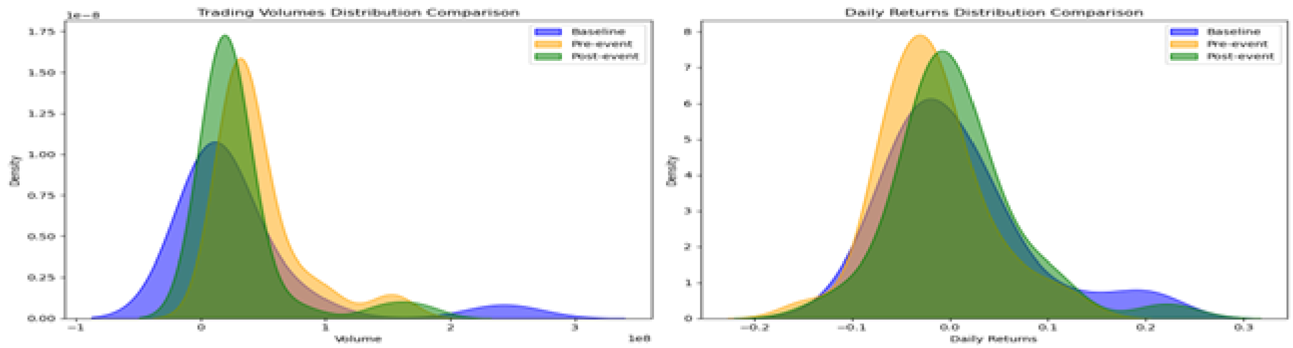


Figure 13: Significant Deviations for trading volumes of CLOV

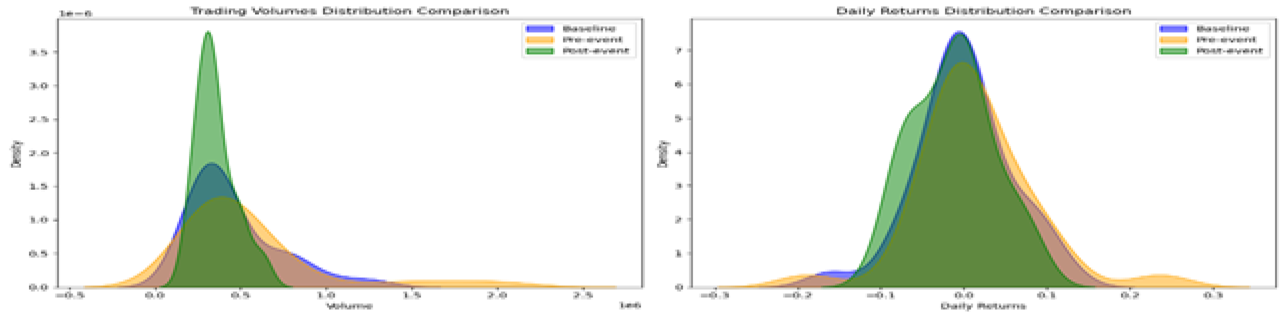


Figure 14: Significant Deviations for trading volumes of ADANIENT.NS

The Z-Score analysis and distribution comparison reveal significant deviations in trading volumes and daily returns for NKLA, CLOV, and ADANIENT.NS post-Hindenburg reports, with $|Z| > 2$ highlighting anomalies and increased volatility, especially for NKLA and CLOV, indicating strong market reactions and substantial shifts in trading behavior.

IV. CONCLUSION

The application of Volume-Weighted Multivariate Kernel Density Estimation (KDE) offers a robust and sophisticated method for detecting potential insider trading activities. By incorporating trading volumes into the KDE framework, this methodology allows for the identification of abnormal price-volume relationships that may signal market manipulation or the misuse of privileged information. Unlike traditional approaches, which may overlook subtle market irregularities, the volume-weighted KDE focuses on periods of heightened trading activity, making it particularly effective in highlighting suspicious market behavior during times of increased volatility. The use of multivariate analysis further enhances the model's precision, as it accounts for the complex interdependencies between various financial variables, such as price movements, volumes, and time. This multivariate approach reduces the risk of false positives by ensuring that only significant deviations from normal trading patterns are flagged. Moreover, the integration of this technique into insider trading detection systems can help in the early identification of price-volume anomalies that may be associated with insider knowledge. This is particularly relevant in modern financial markets, where high-frequency trading and algorithmic strategies can obscure illicit trading patterns. As financial markets continue to evolve, volume-weighted KDE can provide a valuable tool for ensuring market integrity and investor protection. The volume-weighted multivariate KDE offers a more refined lens for analyzing trading data, focusing on periods with abnormal trading volumes and price movements. Its ability to capture and highlight suspicious activities, even in complex market environments, makes it a powerful technique for insider trading detection.

We conclude that "The Hindenburg Research reports triggered significant drops in stock prices due to allegations of corporate misconduct. Nikola faced the most dramatic and lasting decline, driven by legal and regulatory actions, while Adani Enterprises experienced broad scrutiny but managed to regain some market support due to the company's public defence. Clover Health's recovery was aided by retail investor interest, but its reputation still suffered long-term damage. These events illustrate the substantial market impact of investigative reports, particularly when allegations are tied to fraudulent or unethical business practices.

REFERENCES

[1]. L., Kuss, O., Hoyer, A., & Möllenhoff, K. (2023). A non-parametric proportional risk model to assess a treatment effect in time-to-event data. Mathematical Institute, Heinrich Heine University Düsseldorf, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Bielefeld University.

[2]. Acerbi, C., & Tasche, D. (2002). On the coherence of Expected Shortfall. Journal of Banking & Finance, 26(7), 1487-1503.

[3]. Aggarwal, R. K., & Wu, G. (2003). Stock market manipulation — Theory and evidence. Journal of Finance, 58(3), 1407-1430.

[4]. Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), 203-228.

[5]. Biswas, S., & Sen, R. (2018). Kernel based estimation of spectral risk measures. Journal of Statistical Theory and Applications, 17(3), 199-214.

[6]. Bolancé, C., Ayuso, M., & Guillén, M. (2011). Nonparametric approach to analysing operational risk losses. Departament d'Econometria, Estadística i Economia Espanyola, RFA-IREA, University of Barcelona, Spain.

[7]. Brandtner, M. (2019). Spectral risk measures: Properties and limitations: Comment on Dowd, Cotter, and Sorwar. The Journal of Risk, 21(3), 1-7.

[8]. Breunig, R. (2006). An introduction to nonparametric and semi-parametric econometric methods. Australian National University, Sydney.

[9]. Efromovich, S. (2009). Nonparametric regression with missing data: Theory and applications. Statistical Science, 24(3), 452-469.

[10]. Frees, E. W. (2004). Nonparametric estimation of the probability of ruin. North American Actuarial Journal, 8(3), 1-20.

[11]. Haug, J., Hens, T., & Wöhrmann, P. (2013). Risk aversion in the large and in the small. Journal of Economic Behavior & Organization, 86, 83-95.

[12]. Peña, E. A. (2012). Nonparametric statistical methods for complete and censored data. Journal of Nonparametric Statistics, 24(1), 27-46.

[13]. Rizvi, B. A., Belatreche, A., Bouridane, A., & Watson, I. (2017). Detection of stock price manipulation using kernel-based principal component analysis and multivariate density estimation. IEEE Transactions on Information Forensics and Security, 12(5), 1126-1138.

[14]. Wächter, H. P., & Mazzoni, T. (2010). Consistent modeling of risk averse behavior with spectral risk measures (Working paper).

[15]. Wang, B., & Wang, X. (2019). Bandwidth selection for weighted kernel density estimation. Journal of Statistical Computation and Simulation, 89(4), 623-639.