Aalborg University

Asim Hodžić

Mathematics-Economics

Master's Thesis

High Frequency Jump Tests

AALBORG
UNIVERSITY

STUDENT REPORT

AALBORG UNIVERSITY

STUDENT REPORT

Department of Mathematical Sciences
Skjernvej 4A
9220 Aalborg Ø
http://math.aau.dk

Title:
High Frequency Jump Tests

Project:
Master's thesis

Project Period:
Spring semester 2024

Project Group:
Group 4.111C3

Participant:
Asim Hodžić

Supervisor:
J. Eduardo Vera-Valdés

Pages:
23

Date of Completion:
3rd of June

Abstract:

The goal of this thesis is to create efficient tools for detecting jumps in stock prices. To accomplish the goal of creating efficient tools, necessary theory about the statistical tests is introduced. The tools developed are the statistical tests $\mathcal{L}$ and $\hat{G}$ which test jumps in stock price returns and volatility, respectively. Both statistical tests are included in the Julia package `HighFrequencyJumps.jl` which is available on GitHub. Using the developed tools on high frequency stock prices, the thesis shows how the tools should be used. From the analysis on high frequency stock prices, it is shown that the developed tools are efficient in detecting jumps in stock price returns and volatility.

Preface

This Master's Thesis in written in the Spring Semester 2024 by Asim Hodžić, a student studying mathematics-economics at the Department of Mathematical Sciences, Aalborg University.

References are written in the beginning of each chapter and refer to the bibliography. References are written according to the Vancouver method and therefore written as:

[Number, page]

The author of this Master's Thesis would like to acknowledge the supervisor, Associate Professor J. Eduardo Vera-Valdés, for inputs, patience and interest in this Master's Thesis.

Asim Hodžić
ahodzi19@student.aau.dk

Contents

1 | Introduction

The stock market is characterized by continuous fluctuations in stock prices. These fluctuations are not always smooth. The stock market occasionally experiences sudden change in prices. These sudden changes are known as jumps. Jumps in stock prices can result from macroeconomic news, company announcements, to name a few. Being able to understand and detect these jumps is crucial for people in the financial sector since jumps significantly influence the price of a stock.

High frequency data is an important element to be able to capture the intricate behaviors of stock prices. High frequency data provide a detailed view of market activities. This will allow for a precise identification of discontinuities that are missed in lower frequency observations. The ability to be able to detect and analyze a jump using high frequency data is indispensable. Hence, there are some developed statistical tests to detect jumps in high frequency data. Methods for detecting such jumps have been introduced in [1] and [2].

The statistical tests $\mathcal{L}$ and $\hat{G}$ described in [1] and [2], respectively, are powerful tools to detect jumps. These tests utilize high frequency data to detect significant jumps in stock prices. The statistical test $\mathcal{L}$ focuses on identifying jumps by analyzing the stock's return, while the statistical test $\hat{G}$ focuses on identifying jumps by analyzing the volatility. By utilizing the theory of the two tests, the intention of this thesis is to explore the potential of the two tests. A tool for detecting jumps in a stock's return and in volatility will be created to identify jumps in high frequency data. With all this in mind, the main goal of the thesis is answer the following problem statement.

Problem Statement

Can we develop tools to detect jumps in stocks using high frequency data?

2 | Jump Tests

This chapter is based on [1, p. 2535-2543], [2, p. 1-9], [3, p. 2-12], and [4, p. 4-6].

In financial markets, asset prices experience sudden and significant changes known as jumps. Detecting these sudden jumps is crucial when considering various aspects of financial management. Using high frequency data has enhanced the ability to detect such jumps with great precision. Here high frequency data refers to financial data that is observed at short intervals. Utilizing the high frequency data provides a detailed view of price movements which helps with more accurate identification of jumps and allows for the development of non-parametric tests to detect jumps in asset prices. Asset pricing models typically assume that the price paths are continuous. However, jumps could occur due to unexpected market information such as economic announcements or company specific news resulting in paths that are discontinuous.

2.1 Preliminaries

In this section, the intuition behind the detection technique is presented. In addition, the jump detection statistic $\mathcal{L}$ will be mathematically defined later on. The statistic $\mathcal{L}$ tests whether there is a jump at time t_i or not. To get an intuition of the test, envision that asset prices continuously evolve over time in the interval $[0, T]$. The logarithm of the price process S_t for an asset, defined within a probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$, is such that S_t is adapted to a filtration and evolves continuously over time. The price process S_t can be represented as

$$\mathrm{d}\log(S_t) = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, \tag{2.1}$$

where μ_t is the drift, σ_t is the diffusion, and W_t is a standard Brownian motion. Now, suppose at an arbitrary t_i a jump occurs. When there are jumps, the price process of S_t is

$$\mathrm{d}\log(S_t) = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t, \tag{2.2}$$

where J_t is the jump process at time t. The jump process J_t is defined as $J_t = \sum_{i=1}^{N_T} \kappa_{t_i}$, where κ_{t_i} denotes the magnitude of the individual jump at time t_i, and N_T is the counting process that represents the total number of jump up to time T. Now, if a jump occurs at time t_i, the realized return at t_i is expected to be greater than usual. However, there is a problem if the spot volatility at that time is also high. In the absence of jump, given that the spot volatility is high and limited observations at discrete times, the observed realized return may be equivalent to the return because of a jump. To distinguish the scenarios, the returns are standardized using a measure that accounts for the local variation only from the continuous part of the process. The measure that accounts for the local variation is the instantaneous volatility σ_{t_i}, which is incorporated into the jump detection statistic $\mathcal{L}$. Essentially, the statistic $\mathcal{L}$ is the realized return at any given time t_i compared to a consistently estimated instantaneous volatility. However, how can the instantaneous volatility be estimated?

2.2 Estimating Volatility

A frequently employed estimator for the instantaneous volatility is the realized power variation (RPV), also better known as the realized variance. The estimator is defined as

$$\mathrm{RPV} = \sum_{i=1}^{n} r_{t_i}^2, \tag{2.3}$$

where $r_{t_i} = \log(S_{t_i}) - \log(S_{t_{i-1}})$ is the log return at time t_i and n is the number of observations. The RPV is defined by an equally spaced discretization of time, $\Delta t = \frac{T}{n}$, where the interval of interest is $[0, T]$. Furthermore, it is assumed that observations occur every $\Delta t > 0$ periods of time. The interest now lies in observing what happens as $\Delta t \to 0$. The key property for the RPV is given in the following theorem.

Theorem 2.1. Suppose the price process S_t follows (2.2) and that the drift μ_t and diffusion σ_t are independent of the standard Brownian motion W_t. Then, as $\Delta t \to 0$,

$$\mathrm{RPV} = \int_0^T \sigma_s^2 \mathrm{d}s.$$

The theorem states that as the number of observations increase, the RPV converges to the integrated volatility. The integrated volatility measures the total volatility over the time interval $[0, T]$. The integrated volatility offers a thorough assessment of the asset's volatility over that period, capturing the combined impact of changes in instantaneous volatility. However, this frequently employed estimator exhibits inconsistency when jumps occur in a return process [1, p. 2539]. Alternatively, one can use a modified version of RPV. The modified version is the realized bipower variation (RBPV) which is defined as

$$\mathrm{RBPV} = \sum_{i=2}^{n} |r_{t_i}||r_{t_{i+1}}|. \tag{2.4}$$

The key property for the RBPV is given in the following theorem.

Theorem 2.2. Suppose the price process S_t follows (2.2) and the diffusion coefficient is independent of the standard Brownian motion and that the drift is zero. Then, as $\Delta t \to 0$,

$$\text{RBPV} = \int_0^T \sigma_s^2 \mathrm{d}s.$$

The theorem states that the RBPV converges to the integrated volatility as the number of observations increase. However, in contrast to the RPV, the RBPV remains consistent in the presence of jumps in the return process, regardless of the magnitude of the jumps [1, p. 2539], [3, p. 2].

2.3 Detecting Jumps in Returns

In this section, the test that tests whether or not there is a jump at t_i will be presented. Suppose a fixed time horizon T with n number of observations in the interval $[0, T]$ where the distance between two consecutive observations is $\Delta t = \frac{T}{n}$. Now, within the interval $[0, T]$, consider a window of observations of size K. The instantaneous volatility is estimated based on the RBPV using the realized returns within the window of the $K-1$ observations just before testing time t_i. Subsequently, the ratio of this estimated volatility to the subsequent realized return is computed to ascertain the occurrence and magnitude of a jump at time t_i. Mathematically, the test statistic $\mathcal{L}_i$ is defined as follows.

Definition 2.3. The statistic $\mathcal{L}_i$, which at time t_i tests whether there was a jump from t_{i-1} to t_i, is defined as

$$\mathcal{L}_i = \frac{r_{t_i}}{\hat{\sigma}_{t_i}}, \tag{2.5}$$

where r_{t_i} is the log return at time t_i, and

$$\hat{\sigma}_{t_i}^2 = \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} |r_{t_j}||r_{t_{j+1}}|. \tag{2.6}$$

Notice that RBPV is used in the denominator of the statistic (2.5) as the estimation for instantaneous volatility. Using the RBPV as the estimator for the instantaneous volatility in the denominator of the test statistic $\mathcal{L}_i$ ensures that the consistency of estimation remains unaffected by the presence of jumps at earlier times. Hence, the test remains robust in detecting current jumps even in the presence of earlier jumps. Furthermore, to hold on to the benefit of the RBPV, the window size K is selected in such a way that the impact of jumps on the instantaneous volatility estimation vanishes. Additionally, the subsequent theorem states the asymptotic behavior of the jump detection statistic $\mathcal{L}_i$ in instances where there is no jump at time t_i. In addition, suppose that the realized return from time t_{i-1} to t_i is from the diffusion part of the model in (2.1) and (2.2). The asymptotic distribution of the jump detection statistic $\mathcal{L}_i$ is provided in the following theorem.

Theorem 2.4. Let $\mathcal{L}_i$ be defined as (2.5) and $K = O_p(\Delta t^{\alpha})$, where $-1 < -\alpha < -0.5$. Suppose that the process follows either (2.1) or (2.2) and that the drift and diffusion coefficients remain relatively constant within a short time span. Let $\bar{A}_n$ be the set of $i \in \{1, 2, \ldots, n\}$ such that there is no jump in $(t_{i-1}, t_i]$. Then, as $\Delta t \to \infty$,

$$\sup_{i \in \bar{A}_n} |\mathcal{L}_i - \hat{\mathcal{L}}_i| = O_p\left(\Delta t^{\frac{3}{2}-\delta+\alpha-\epsilon}\right),$$

where δ satisfies $0 < \delta < \frac{3}{2} + \alpha$ and

$$\hat{\mathcal{L}}_i = \frac{U_i}{c}.$$

Here $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$ is a standard normal variable and $c = \frac{\sqrt{2}}{\sqrt{\pi}}$ is a constant.

The above theorem states that the statistic $\mathcal{L}_i$ approximately follows that same distribution as $\hat{\mathcal{L}}_i$ when there is no jump at t_i. As Theorem 2.4 states, $\hat{\mathcal{L}}_i \to \mathcal{N}(0, 1)$ since U_i is a standard normal random variable. However, how does the test statistic $\mathcal{L}_i$ behave upon arrival of jumps. The following result shows that as the sampling interval $\Delta t \to 0$, the test statistic $\mathcal{L}_i$ becomes so large such that it is possible to detect a jump.

Theorem 2.5. Let $\mathcal{L}_i$ be defined as (2.5) and $K = O_p(\Delta t^{\alpha})$, where $-1 < -\alpha < -0.5$. Suppose that the process follows (2.2) and that the drift and diffusion coefficients remain relatively constant within a short time span. Suppose there is a jump at any time $\tau \in (t_{i-1}, t_i]$. Then

$$\mathcal{L}_i \approx \frac{U_i}{c} + \frac{\kappa_\tau}{c\sigma\sqrt{\Delta t}},$$

where κ_τ is the actual jump size at jump time τ. Hence, $\mathcal{L}_i \to \infty$ as $\Delta t \to 0$. If there is no jump at any time $\tau \in (t_{i-1}, t_i]$, $\mathcal{L}_i$ has the asymptotic behavior as in Theorem 2.5.

When considering the RBPV, the benefit of employing it as an instantaneous volatility estimator in the denominator of the test statistic is that the presence of jumps at earlier times does not affect the consistency of the estimation. Therefore, the test is robust to earlier jumps in detecting current jumps. Furthermore, to hold onto the benefit of the RBPV, the window size K is selected in such a way that the impact of jumps on the instantaneous volatility estimation vanishes. In addition, the window size K clearly has to be smaller than the number of observations ($nobs$) n. The selection of sampling frequency Δt will dictate the size of the window K. Typically, the number of observations per day are calculated as $\Delta t = \frac{1}{252 \times nobs}$. Hence, integers between $\sqrt{252 \times nobs}$ and $252 \times nobs$ are candidates for the window size K. However, in [1, page 2542] they suggest some optimal window sizes K. They are as follows: one week (7), one day (16), one hours (78), 30 minutes (110), 15 minutes (156), and 5 minutes (270).

Now, the rejection region for the test statistic will be introduced. Taking a look at what is described in Theorem 2.4 and Theorem 2.5, the test statistic presents different behavior depending on the existence of jumps at testing times; if there is not a jump, the test statistic approximately follows a normal distribution, however, if there is a jump, the test statistic becomes very large. To ascertain a reasonable rejection region, the question revolves around how large the test statistic can be in the absence of a jump. Thus, the initial focus is taking a look at the asymptotic distribution of maximums of the test statistics when there is no jumps at any time in $(t_{i-1}, t_i]$. Such a distribution informs the selection of relevant threshold for the test to distinguish when there is a jump at a testing time. The following lemma states the limiting distribution of the maximums.

Lemma 2.6. Let $\mathcal{L}_i$, K, and $\bar{A}_n$ be as in Theorem 2.4. Then, as $\Delta t \to 0$,

$$\frac{\max_{i \in \bar{A}_n} |\mathcal{L}_i| - C_n}{S_n} \to \xi,$$

where ξ has cumulative distribution function $\mathbb{P}(\xi \leq \beta^*) = \exp(-\exp(-\beta^*))$,

$$C_n = \frac{\sqrt{2\log(n)}}{\sqrt{\frac{2}{\pi}}} - \frac{\log(\pi) + \log(\log(n))}{2\sqrt{\frac{4}{\pi}\log(n)}} \quad \text{and} \quad S_n = \frac{1}{\sqrt{\frac{4}{\pi}\log(n)}},$$

where n is number of observations.

Note that the the normalizing constants C_n, S_n have been demonstrated to hold [5]. To utilize the above result for selecting a rejection region, start by choosing a significance level of, for instance, $\alpha = 5\%$. Then, the threshold for $\frac{|\mathcal{L}_i| - C_n}{S_n}$ is β^*, such that $\mathbb{P}(\xi \leq \beta^*) = 0.95$. Here β^* is calculated to be 2.9702. Hence, the hypothesis of no jump at time t_i is rejected whenever $\frac{|\mathcal{L}_i| - C_n}{S_n} > 2.9702$. Additionally, by observing the sign of $\mathcal{L}_i$ it is possible to determine the direction of the jump, which is crucial.

2.4 Detecting Jumps in Volatility

In addition to the jump detection statistic $\mathcal{L}$, a comparison to one of the first tests for detecting jumps in high frequency will be made. The test in question is the Barndorff-Nielsen and Shephard test (BNS test), specifically the linear jump statistic $\hat{G}$. The test compares the RBPV with RPV, respectively equation (2.4) and (2.3). However, to be able to construct a computable linear jump test, the integrated quarticity (IQ) $\int_0^T \sigma_s^4 \mathrm{d}s$ has to be estimated. The IQ can be consistently estimated using the realized quadpower variation (RQPV) which is defined as

$$\text{RQPV} = \sum_{i=4}^{n} |r_{t_i}||r_{t_{i+1}}||r_{t_{i+2}}||r_{t_{i+3}}|. \tag{2.7}$$

Now that it is possible to consistently estimate the IQ, the statistic is defined as

$$\hat{G} = \frac{\frac{\pi}{2}\text{RBPV}_{t_i} - \text{RPV}_{t_i}}{\sqrt{\frac{1}{n}\frac{\pi^2}{4}\vartheta\text{RQPV}_{t_i}}}. \tag{2.8}$$

Here $\vartheta = \frac{\pi^2}{4} + \pi - 5$ is a coefficient that is related to the IQ. Furthermore, Theorem 1 in [2, p. 6] states the asymptotic distribution. The theorem states that the test is $\mathcal{N}(0,1)$ under the null hypothesis that there are no jumps. However, when (2.8) becomes very large or small, the null hypothesis is rejected.

3 | Using Jump Tests

In this chapter, the relevant theory that has been introduced in Chapter 2 has been coded in Julia Programming Language (henceforth referred to as Julia) and will be thoroughly explained. The objective is to explain how one can use these tests or tools in practice. The jump tests are included in the `HighFrequencyJumps.jl` package and can be found on GitHub.

3.1 Jumps in Returns

The $\mathcal{L}$ statistic, which is defined in Definition 2.3, tests whether there is a jump from t_{i-1} to t_i. The test has been coded in Julia as the function `LMJumpTest(S, K, α)`. As can be seen, the function has three inputs: `S`, `K`, α. Here `S` are the prices, `K` is the window size, and α is the significance level. In addition to Definition 2.3, Lemma 2.6 is implemented into the function `LMJumpTest(S, K, α)` to check the null hypothesis.

The function `LMJumpTest(S, K, α)` starts by finding the length of the input `S`, which are the prices of a stock for instance. We do this since we need to calculate all the returns in (2.5). Then we initialize a vector of zeros, where we will store the computed $\mathcal{L}$ statistic. Hence, the function `LMJumpTest(S, K, α)` now starts by looping from the i'th return. This loop goes from $K+1$ up to n. The reason it starts from $K+1$ and not just K is because when we calculate the RBPV in (2.6) it will start from S_0, which is the initial price in theory. However, the initial price is S_1 in Julia since a loop can not start from zero. Nevertheless, inside the loop is another loop. The inside loop starts by initializing a number where the loop calculates the RBPV from $i-K+2$ up to $i-1$. After it has calculated the RBPV in (2.6), the RBPV is divided by $K-2$ and then assigned to $\hat{\sigma}^2$. The last step is to divide the i'th log return by the standard deviation. The loop will go through all log returns and calculate the $\mathcal{L}$ statistic such that we can access them when we want check the null hypothesis. Note that we check if the RBPV is greater than zero such that we do not divide by zero. If it is zero, the $\mathcal{L}$ statistic is set to zero.

Now that the $\mathcal{L}$ statistic has been calculated, we want to test whether to reject the hypothesis of no jump at time t_i or not. This is done by looking at Lemma 2.6. We start by defining the normalizing constants C_n, S_n and β^*. Here β^* is defined such that it is possible to choose any significance level α, which is an input in the function `LMJumpTest(S, K, α)`. Nevertheless, we initialize a vector containing a boolean, which is a true or false statement. A loop then goes through all the $\mathcal{L}$ values to test whether there is a jump or not. Finally, the function `LMJumpTest(S, K, α)` returns the vector of $\mathcal{L}$ values and a vector $\mathcal{H}_0$ containing the statements "true" (1) and "false" (0) where "true" is a jump.

3.2 Jumps in Volatility

The $\hat{G}$ statistic, which is defined in (2.8), tests whether there is a jump in volatility at time t_i. The test has been coded in Julia as the function `BNSJumpTest(S)`. As can be seen, the function has one input: `S`. Here `S` are the prices again.

The function `BNSJumpTest(S)` starts by finding the length of the input `S`. We do this since we need to calculate the RPV, RBPV, and RQPV in (2.3), (2.4), and (2.7), respectively, which all go up to n. Then the coefficient related to the IQ is defined. In addition, we initialize three vectors of zeros, where we will store the calculated RPV, RBPV, and RQPV values. The function then continuous to calculate all the values of RPV, RBPV, and RQPV for every i. Once they have been calculated, we initialize a vector of zeros, where we will store the $\hat{G}$ values. These are calculated using (2.8). Note that the the sum to calculate (2.8) starts from five and not four as the sum suggests in (2.7). This is the same reason as for the other test where the initial price is S_1 in Julia since a loop can not start from zero. In addition, it is taken into account that the $\hat{G}$ values are computed for i starting from five to n, which ensures it uses the appropriately indexed values from RPV, RBPV, and RQPV. This consistency in indexing ensures that all calculations are aligned with the same sequence of prices. Finally the function `BNSJumpTest(S)` returns the $\hat{G}$ values. Note that we check if the RQPV is greater than zero such that we do not divide by zero. If it is zero, the $\hat{G}$-statistic is set to zero.

3.3 Applying Jumps in Returns

Now we dive into how it is intended to use the Julia function `LMJumpTest(S, K, α)`. First of all, we need some prices, a window size, and a significance level. Once those are established, we can use the function `LMJumpTest(S, K, α)`. In what follows is the intended use of the function.

```
1 ℒ, ℋ0 = LMJumpTest(S, K, α)
2 histogram(ℒ[K+1:length(ℒ)], normalize=:pdf)
3 num_true = count(==(true), ℋ0)
4 index_true = findall(==(true), ℋ0)
```

As explained before, the `LMJumpTest(S, K, α)` function returns a vector of $\mathcal{L}$ values and a vector $\mathcal{H}_0$ containing the statements "true" and "false". Hence, do as in line 1 to extract the $\mathcal{L}$ values and $\mathcal{H}_0$ statements. Once they have been extracted, we start by checking how the $\mathcal{L}$ values are distributed. This is what line 2 does. Note that we use the values from $K + 1$ since the first K values are used to calculate the first value of $\mathcal{L}$ and are therefore zero. Nevertheless, once it has been shown how the distribution of the $\mathcal{L}$ values is, we can dive in and look at how many jumps the function has actually detected. This is what line 3 does. Essentially, this just says we detect X amount of jumps. Line 4 will find at what index the jump is. These indices can now be used to look at what point in time did a jump occur, and can be analyzed to see why the jump might have occurred.

3.4 Applying Jumps in Volatility

The other function we dive into to see how it is intended to use is the Julia function `BNSJumpTest(S)`. All we need in this function are some prices. Once those are established, we can use the function `BNSJumpTest(S)`. In what follows is the intended use of the function.

```
1 Ĝ = BNSJumpTest(S)
2 histogram(Ĝ[5:length(Ĝ)])
3 min5 = partialsort(Ĝ, length(Ĝ)-4:length(Ĝ))
4 max5 = partialsort(Ĝ, 1:5)
```

As explained before, the `BNSJumpTest(S)` returns a vector of $\hat{G}$ values. Hence, do as in line 1 to extract the $\hat{G}$ values. Once they have been extracted, we start by checking how the $\hat{G}$ values are distributed. This is what line 2 will check. Line 3 and 4 will give the smallest and biggest $\hat{G}$ values, respectively. These will be used to find at what point in time did a jump occur, and can be further analyzed to see why the jump might have occurred.

4 | Jump Tests on Stock Prices

In this chapter, the tests will be used on high frequency data. In particular, the tests will be used on Vestas Wind Systems (VWS) stock prices as well as Tesla, Inc. (TSLA) stock prices. The stock prices of VWS are high frequency data down to 1 second while the TSLA stock prices are observed every 5 minutes. The stock prices of VWS are from 16/03/2023 to 11/04/2023 and the stock prices of TSLA are from 29/04/2024 to 20/05/2024. The stock prices are from when their respective market is open. For VWS, the market is open from 09:00 to 16:55 while the market for TSLA is open from 09:30 to 16:00. The prices of the two stocks are presented in the following figures.

FIGURE 4.1: VWS's price movement.

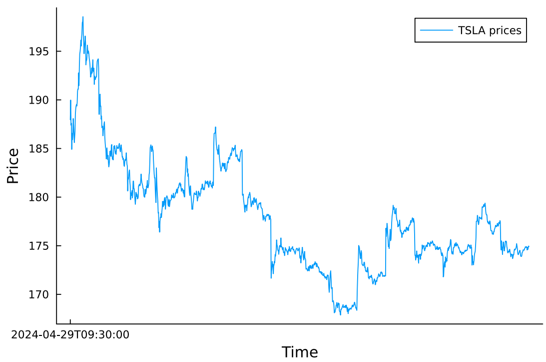

FIGURE 4.2: TSLA's price movement.

The number of observations for VWS and TSLA are 34178 and 1264, respectively.

4.1 Jump Tests on TSLA prices

We start by using the Julia function `LMJumpTest(S, K, α)` on TSLA prices. Here we have to choose a window size and a significance level. The choice of window size will be 77 which means we use prices from a whole trading day to calculate RBPV. This is a suggestion they have in [4, p. 5], where they use data that is observed every 5 minutes. The choice of significance level is 1% since we want to look at the jumps that are extreme. First we will take a look at how the calculated $\mathcal{L}$ values are distributed.

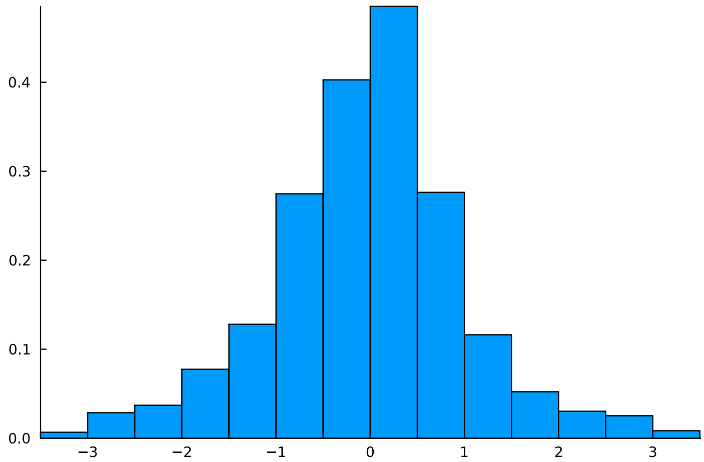

Figure 4.3: Distribution of the calculated $\mathcal{L}$ values.

As can be observed in Figure 4.3, it seems that the $\mathcal{L}$ values are normally distributed around 0. The mean of the $\mathcal{L}$ values is approximately -0.04 while the standard deviation is approximately 1.70 which could indicate that there might be jumps in TSLA prices. Indeed, the test statistic $\mathcal{L}$ detects 14 jumps in the price series. Out of the 14 jumps, 12 of them are from when the market opens with the remaining two being from when the market has been open for a while. Relevant information about the two jumps can be found in the following table.

Time of detection	**Return at time t_i**	**$\mathcal{L}$ at time t_i**
10:25:00 on 10th of May	-1.47	-5.76
10:30:00 on 17th of May	1.73	6.41

Table 4.1: Information about jumps.

As can be seen in Table 4.1, we have two instances where the price jumps a significant amount up and down. Now it will be analyzed what could have caused the price to jump up and down. On the 10th of May, there were protests at a Tesla factory in Germany. They were demonstrating against the environmental impact of the construction of the new factory. The incident involved clashes with police which lead to arrests[1]. This negative publicity surrounding the protests and potential disruptions to operations could have contributed to a decrease in investor confidence. This could potentially have led to the fall in TSLA's stock price the 10th of May. On the other hand, there were some other news on the 17th of May where there were news about job reduction at Tesla. Around 600 positions were eliminated in California[2]. Considering these adjustments, the increase

[1]Article can be found on cnbc.com.
[2]Article can be found on cnbc.com.

in TSLA's stock price could be associated to factors such as cost savings. Investors might see these changes as good, thinking that Tesla will run better and be more profitable in the future.

Now we will take a look at the other test using the Julia function `BNSJumpTest(S)` on TSLA price. Here we just need the prices to calculate the $\hat{G}$ values. First we will take a look at how the calculated $\hat{G}$ values are distributed. Note that if the denominator is zero in (2.8), the function `BNSJumpTest(S)` will return a zero. These will be removed when looking at how the $\hat{G}$ values are distributed.

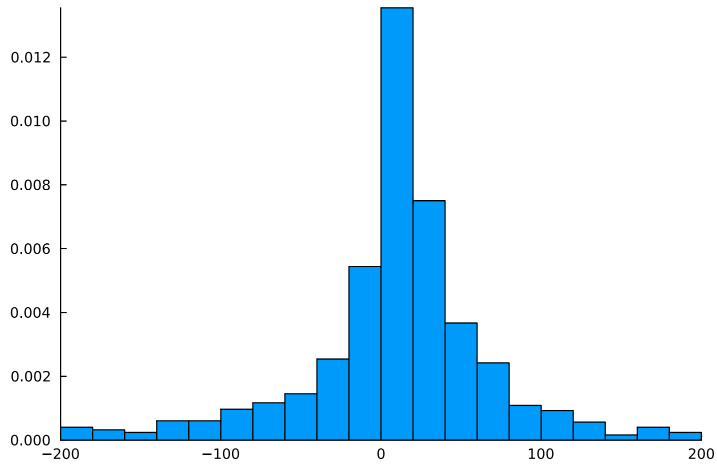

Figure 4.4: Distribution of the calculated $\hat{G}$ values.

As can be observed in Figure 4.4, it seems that the $\hat{G}$ values are normally distributed. However, with a mean of approximately -194.14 and a standard deviation of approximately 4350.03 could indicate that there might be jumps in volatility. As described, we will be looking at the five biggest and smallest $\hat{G}$ values and then analyze at which point in time the jump occurred. Of those ten values, five of them are from when the market opens. The remaining five are from when the market has been open for a while. Relevant information about the five jumps in volatility can be found in the following table.

Time of detection	$\hat{G}$ **at time** t_i
15:05:00 on 29th of April	914.34
15:40:00 on 29th of April	947.38
15:40:00 on 16th of May	1440.61
15:00:00 on 20th of May	-127673.00
15:05:00 on 20th of May	25382.16

Table 4.2: Information about jumps.

As can be seen in Table 4.2, we have three different days where the jump in volatility is significant. Now it will be analyzed what could have caused the jump in volatility. Note that it does not detect the same days as in Table 4.1. On the 29th of April, TSLA's stock price surged following the company's significant progress in gaining approval for its Full Self-Driving technology in China[3]. News on 16th of May could have caused a jump in volatility due shareholders urging a vote against Elon Mush's pay package, which could have a negative impact on the company[4]. On the 20th of May could have caused a jump in volatility due to Tesla loosing a legal appeal in Sweden concerning a labor strike. Swedish authorities blocked the issuance of new license plates for Tesla vehicles during the strike[5].

4.2 Jump Tests on VWS prices

As for TSLA prices, we start by using the Julia function `LMJumpTest(S, K, α)` on VWS prices. Here we have to choose a window size and a significance level again. The choice of significance level is again 1% since we want to look at jumps that are extreme. However, the choice of window size will be 2135 following the same logic as before - we use prices from a whole trading day to calculated RBPV. In the period of the observed VWS stock prices, there are 16 trading days. This means there are an average of $\frac{34178}{16} = 2136.125$ observations per day. Nevertheless, we will first take a look at how the calculated $\mathcal{L}$ values are distributed.

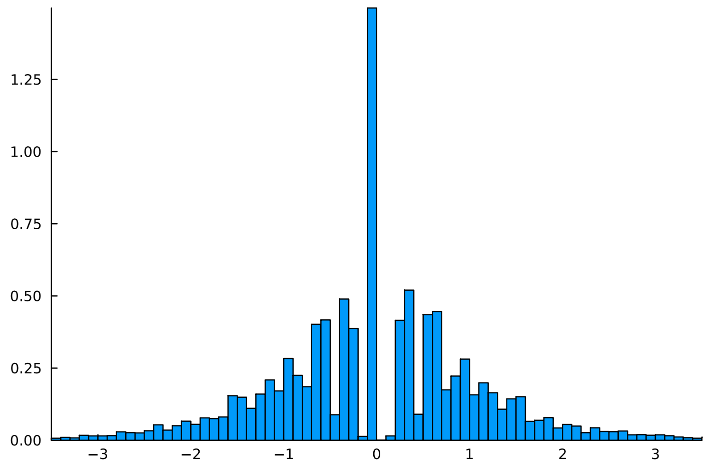

Figure 4.5: Distribution of the calculated $\mathcal{L}$ values.

As can be observed in Figure 4.5, it seems that the $\mathcal{L}$ values are somewhat normally distributed. The mean is approximately zero while the standard deviation is approximately

[3]Article can be found on cnbc.com.
[4]Article can be found on bloomberg.com.
[5]Article can be found on bloomberg.com.

1.50. This could indicate that there might be jumps in VWS prices. Indeed, the test statistic $\mathcal{L}$ detects 62 jumps. Out of those 62 jumps, 56 of them are from when the market opens or close to the other remaining jumps from when the market has been open for a while. Relevant information about the remaining six jumps can can be found the in following table.

Time of detection	**Return at time t_i**	**$\mathcal{L}$ at time t_i**
15:29:16 on 17th of March	0.44	8.30
11:43:36 on 21st of March	0.65	10.24
12:01:37 on 24th of March	-0.42	-7.01
10:01:31 on 27th of March	0.60	8.92
15:12:00 on 4th of April	-0.5	-6.65
16:23:39 on 11th of April	0.40	7.63

TABLE 4.3: Information about jumps.

As can be seen in Table 4.3, we have six instances where the price jumps a significant amount up and down. Now it will be analyzed what could have caused the price to jump up and down. On the 17th of March, the VWS stock price likely increased due to securing a 50 MW order in Finland and Germany[6]. On the 21st of March there were some news about VWS buying ST3 Offshore[7]. This could have driven the VWS stock price up since investors see positively on the VWS expansion in the offshore wind sector. Lastly, on the 11th of April there were news about VWS figuring out how to recycle old wind turbines, which is ground breaking news for the company[8]. This could have resulted in a rise in stock price since investors may see benefits when considering cost savings. For the other detected jumps, no related news on the price movements were found. However, taking a look at Figure 4.1, a big jump can be noticed. The jump is detected on the 30th of March when the market has to opened. The return at that time is 6.78 with an $\mathcal{L}$ value of approximately 118. This is a big return and jump which is due to securing a two huge orders of 1.310 MW in Brazil and 162 MW in Argentina[9]. Generally speaking, for VWS stock prices, when there are news about huge orders, their prices tend to increase significantly, which should be taken into account.

Now we will use the other test using the Julia function `BNSJumpTest(S)` on VWS prices. Here we just need the prices to calculate the $\hat{G}$ values. First we will take a look at how the calculated $\hat{G}$ values are distributed. Note that if the denominator is zero in (2.8), the function will return a zero. As for TSLA, these will be removed when looking at how the $\hat{G}$ values are distributed. For the VWS, the denominator was 0 in 15427 instances. These are removed to see how the $\hat{G}$ values are distributed.

[6] Wind Turbine Orders can be found on vestas.com.
[7] Article can be found on renewablesnow.com.
[8] Article can be found on berlingske.dk
[9] Wind Turbine Orders can be found on vestas.com.

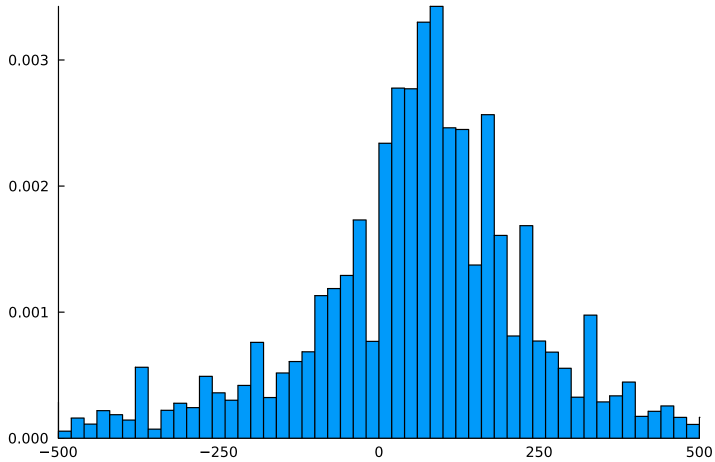

FIGURE 4.6: Distribution of the calculated $\hat{G}$ values.

As can be observed in Figure 4.6, it seems that the $\hat{G}$ values are heavily skewed to right. Furthermore, is has a mean of approximately -2.17 and a standard deviation of approximately 523.67. These findings could indicate that there might be jumps in volatility. As for TSLA, we will be looking at the five biggest and smallest $\hat{G}$ values and then analyze at which point in time the jump occurred. Of those ten values, nine of them are due to the market has to opened. The last detected jump in volatility is detected at 24th of March where it was not possible to find related news. However, when looking at the smallest (-34923.35) and biggest (2785.90) value of $\hat{G}$, we detect jump in volatility at the 27th of March and 11th of April. Both are detected at when the market has just opened and could be the same reason as described before.

5 | Discussion

It has been shown that the developed tools for detecting jumps in return and volatility work on high frequency data. However, when applying the tools on high frequency data, some complications may occur. When applying the Julia function `BNSJumpTest(S)` on VWS stock prices, we ended up dividing by zero in the denominator in (2.8) 15427 times. By further investigation, the dataset contained prices, where two consecutive prices could be identical. This means if this happens at a given time, the return will be zero. By taking a look at how the RQPV is defined in (2.7), one sees that if one of the returns give zero, the whole sum will give zero. This will cause division by zero in (2.8), which is not possible. Take into account these findings to ensure that when applying the `BNSJumpTest(S)` on high-frequency data, situations where there are two consecutive identical prices, are properly addressed. This also applies for Julia function `LMJumpTest(S, K,` α`)`. However, note that the numerator can be zero in (2.5) which would just mean there are no change in price at that specific point in time. Nonetheless, applying `BNSJumpTest(S)` on TSLA stock prices did not cause any problems.

When applying the Julia function `LMJumpTest(S, K,` α`)` on stock prices, the choice of an appropriate window size turned out to be challenging. Considering a window size of 270, which is suggested in section 2.3, the function `LMJumpTest(S, K,` α`)` detected nine jumps when applied on TSLA stock prices. This is five less than when using a window size of 77. Upon further examination, the nine jumps were already detected when using the window size of 77. In addition, it only detects jumps when the market opens. However, what happens when a window size of 39, which is half a trading, is used? When a window size of 39 is used, the function `LMJumpTest(S, K,` α`)` detected 17 jumps. However, all of them are from when the market has just opened. In addition, it also detects the same jumps when the window size is 77. This suggests that using observations from an entire trading as the window size could be optimal. The same applies for the VWS stock prices.

6 Conclusion

In order to be able to develop tools to detect jumps in stocks using high frequency data, necessary theory about the statistical jump tests has been introduced. The statistical tests $\mathcal{L}$ and $\hat{G}$ are powerful tools to detect jumps in high frequency data. The $\mathcal{L}$ statistic focuses on identifying jumps by analyzing the return while the $\hat{G}$ statistic focuses on identifying jumps by analyzing the volatility.

When using the developed tools `LMJumpTest(S, K, α)` and `BNSJumpTest(S)` on high frequency data, the results indicate that the developed tools can detect jumps using high frequency data. The `LMJumpTest(S, K, α)` tool has its adjustable parameters that offer some flexibility in identifying jumps in stock prices using high frequency data. Meanwhile, `BNSJumpTest(S)` has also proven to be a tool usable for identifying jumps.

Their application to VWS and TSLA stock prices reveal the accuracy of the tools. Both tools, `LMJumpTest(S, K, α)` and `BNSJumpTest(S)`, show at which point in time the jumps occur. Typically, the jumps detected using the tools occurred when the market has just opened. This was not a surprise. However, the tools did find jumps when the market has been open for quite some time. When analyzing these jumps, we found some relevant news. This means, when significant news events occur - good or bad - the developed tools can help investors identify which events are likely to impact the market and thereby determine the appropriate actions to take. However, when applying the `LMJumpTest(S, K, α)` tool, one should be aware of how to choose the window size `K`. What seemed to work when analyzing VWS and TSLA stock prices was using observations from a whole day. This might not be the case in other applications.

The tools can be found in the Julia package `HighFrequencyJumps.jl` on GitHub, which provide both the tools for detecting and analyzing jumps on high frequency data. The tools are user-friendly and efficient allowing others to easily integrate the developed tools into their own analysis of for instance stock prices.

Bibliography

[1] Lee SS, Mykland PA. Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. The Review of Financial Studies. 2007;.

[2] Barndorff-Nielsen OE, Shephard N. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. Journal of Financial Econometrics. 2006;.

[3] Barndorff-Nielsen OE, Shephard N. Power and Bipower Variation with Stochastic Volatility and Jumps. Journal of Financial Econometrics. 2004;.

[4] Kvist O, Vera-Valdés JE. The Effect of CEO Public Behaviour on the Company's Valuation: The Case of Tesla and Elon Musk. Department of Mathematical Sciences, Aalborg University. 2024;.

[5] Nunes JPV, Ruas JP. A note on the Gumbel convergence for the Lee and Mykland jump tests. Finance Reseach Letters. 2024;.