Analyzing and Applying Existing and New Jump Detection Methods for Intraday Stock Data

W. Warren Davis
WWD2@DUKE.EDU

Professor George Tauchen, Faculty Advisor

Honors submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University; Durham, North Carolina
2008

Acknowledgements

Many people are responsible for the work that was completed in this research. I would like to thank the members of the Duke University Econometrics and Finance Lunch Group for their feedback and suggestions. In particular, I would like to thank the members of the Honors Senior Finance Workshop group, George Tauchen, Tim Bollerslev, Bjorn Eraker, Tzuo Law, Peter Van Tassel, and Andrey Fradkin.

Abstract

This paper attempts to explore two recent statistics used to identify jumps in stock prices, as well as to propose a modification to one of the statistics to increase its accuracy by adding a second stage with a different estimator of local volatility. After identifying potential jump days, a study of Bristol-Myers Squibb Co. stock was performed, identifying the types of company-specific events that occurred on these days that seemed to cause jumps in the price. Also, the new proposed statistic was found to be more accurate by a using method of changing the significance levels used in each stage, as well as in samples with an extremely high jump frequency.

For many years, financial economists assumed that security returns followed a continuous random walk, which could be described by a normal distribution. Many key models used this assumption, such as the Black-Scholes model for pricing options (Black and Scholes, 1973). However, as these returns were examined further, it seemed clear that they were not normally distributed; the distributions of the returns seemed to have "fat tails," meaning large moves in prices were more common than a normal distribution would predict. Also, statisticians found that there were in fact jumps, very large either positive or negative changes in prices (Andersen, Bollerslev, and Diebold, 2002), which contradicted models such as the Black-Scholes. Recently, financial economists have begun to attempt to identify and quantify these random jumps in returns by using high-frequency stock data. One of the major statistics used to try to identify such jumps, discussed in Barndorff-Nielsen and Shephard (2004) and Huang and Tauchen (2005), claims to identify which trading days contain at least one jump. While the results from this statistic are well documented, other economists sought to overcome the limitation that the statistic cannot identify multiple jumps in a day, or the exact time of a jump; it merely attempts to determine whether or not at least one jump was present during the trading day.

Two of these economists were Lee and Mykland (2006), who propose a statistic that not only determines whether or not at least one jump was present during the trading day, but would also identify multiple jumps in a day, as well as their exact times. The two economists used a statistic that attempts to label the specific returns where jumps occurred by constructing a ratio of the return to a measure of local volatility that is created by using a trailing average of the Bipower Variation (a statistic that will be defined in depth later in the paper). By using this method, large jumps in the stock price could be identified.

This paper attempts to determine if Lee-Mykland's statistic might be modified so that it becomes more efficient. While the Bipower Variation is robust to jumps, Huang and Tauchen (2005) document that the Realized Variance (a statistic which will also be defined later) is a more efficient estimator, although it is not robust to jumps. Therefore, it is hypothesized that a more efficient estimator can be used, namely the Realized Variance, and a new statistic is proposed.

This new statistic is similar to Lee and Mykland's but conducted twice, slightly modifying the statistic during the second step. First, the original Lee-Mykland statistic is calculated, identifying returns as potential jumps. The second iteration then uses the Realized Variance to estimate local volatility; however, since it is not robust to jumps, all jumps flagged by the Lee-Mykland statistic in the first stage are removed through processes that will be described later, and the Realized Variance is used to compute the new estimate of the local variance using this smoother, jump-free set of returns. Finally, the t-statistic is re-computed, using the original returns in the numerator and the new jump-free estimator of local variance in the denominator. It is hypothesized that this new method creates a statistic that is a more accurate jump detector, since it might slightly smooth out the estimated local volatility and cause jump returns to stand out. It is proposed that if the new statistic is able to detect jumps more accurately, then it would be a useful addition to Lee and Mykland's statistic; on the other hand, if it is not significantly more accurate in flagging jumps, then it would uphold and strengthen Lee and Mykland's original statistic. This paper will describe the process that led to the proposed statistic, and the results of testing the accuracy of both the Lee-Mykland statistic and the new two-stage approach in correctly flagging jumps in data simulated using various models.

First, the paper will begin with a brief discussion of the high-frequency stock data used throughout the research. Next, it will outline the methods and models used to conduct the three jump detection tests and outline the statistics presented by Barndorff-Nielsen and Shephard (BNS), the Lee-Mykland statistic (L-M), and the modified, two-stage process described above. Next, the results of applying these statistics to actual high-frequency stock data will be presented. In addition, this section will give an overview of the results obtained when each day flagged by the Barndorff-Nielsen and Shephard statistic as containing a jump was examined using the Factiva news service, attempting to identify any company-specific events that might have caused a jump in stock prices. The final section will present two price models used to create simulated series of stock returns, and the results when both the original Lee-Mykland statistic and the modified two-stage statistic are applied to these returns.

I. Data

This research examines high-frequency stock data from the New York Stock Exchange. Specifically, the research was primarily performed on the stock of Bristol-Myers Squibb Co (ticker symbol BMY). The actual data sets, containing 30-second returns from all full trading days from January 2001 through December 2005, were acquired from the Trade and Quote Database (TAQ), which was obtained through Wharton Research Data Services. While this paper will provide a brief overview, a more thorough description of this data can be found in Law (2007).

While these data are quite reliable, there are some entry errors in some of the data sets; therefore, the data needs to be cleaned up. Even though only the 40 most actively traded stocks on the NYSE are even considered to be analyzed since high liquidity is desired, there are still some erroneous trades and data entry errors that need to be corrected. Tzuo Law performed the initial work, using an adapted version of the previous tick method from Dacorogna, Gencay, Muller, Olsen, and Picter (2001). This method excludes the first five minutes of the trading day, so that trading is more uniform. Therefore, there are 771 observations for each of the 1241 days examined, going from 9:35 am to 4:00 pm. However, since market microstructure noise increases as the sampling intervals become smaller, 5-minute returns were examined throughout the research to lessen its effect. Finally, two methods are used to manually clean up the data. First, whenever there are two consecutive returns over 1.5% in opposite directions, both returns are set equal to zero. This is used because it was decided that two such offsetting returns are most likely the result of a data entry error, as such a move would be very unlikely to occur within normal trading conditions. Also, in some cases, simple manual inspection is used. If there seemed to be a spurious return whose magnitude made no sense within a series of returns, it is also set equal to zero. By using these two methodologies to clean up the data set, it is believed that most errors within the data sets can be eliminated.

II. Methods Used in Statistics

This section will seek to provide a step-by-step explanation of how each statistic is constructed and conducted. First, it will describe the statistic found in Barndorff-Nielsen and Shephard (2004) and Huang and Tauchen (2005), Lee and Mykland's statistic, and finally the modified two-stage approach.

A. The Statistic Presented by Barndorff-Nielsen and Shephard to Identify Jump Days

The analysis is performed under the assumption that the log-price *p(t)* is defined in continuous time as follows (Huang and Tauchen 2005):

$$dp(t) = \mu(t)dt + \sigma(t)dw(t) + d\mathcal{L}_J(t)$$

This model consists of a drift term added to a standard Brownian motion multiplied by the instantaneous volatility. The final term shows a pure jump Levy process, with increments $Lj(t) - Lj(s) = \Sigma_{s\leq\tau\leq t}\, \kappa(\tau)$, where $\kappa(\tau)$ is the jump size. The specific Levy process examined is a Compound Poisson-Process (CPP), where jump intensity is constant, and jump size is independently identically distributed.

The statistic presented by Barndorff-Nielsen and Shephard and later examined by Huang and Tauchen utilizes several statistics, as presented below. First, the return $r_{t,j}$ is defined as simply the difference in each consecutive log-price, as defined above. Next, the Realized Variance, as presented in Andersen, Bollerslev, and Diebold (2002), is defined as

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2,$$

and the Bipower Variation is defined as

$$BV_t = \mu_1^{-2}\left(\frac{M}{M-1}\right)\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}| = \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}|$$

.

What is very important about these two estimators of integrated variance is that,

$$\lim_{M\to\infty} RV_t = \int_{t-1}^{t}\sigma^2(s)ds + \sum_{j=1}^{N_t}\kappa_{t,j}^2,$$

and, according to Barndorff-Nielsen and Shephard (2004a), together with Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2005) and Barndorff-Nielsen, Graversen, Jacod, and Shephard (2005), under reasonable assumptions,

$$\lim_{M\to\infty} BV_t = \int_{t-1}^{t} \sigma^2(s)ds.$$

These limits show that the Realized Variance is a consistent estimator of integrated variance summed with the jump contribution, while Bipower Variation is a consistent estimator of the integrated variance, regardless of the presence of jumps. Therefore, by using the difference of these two limits, a consistent estimator of the jump contribution, RV_t - BV_t , can be used, since

$$\lim_{M\to\infty} RV_t - BV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{j=1}^{N_t} \kappa_{t,j}^2 - \int_{t-1}^{t} \sigma^2(s)ds = \sum_{j=1}^{N_t} \kappa_{t,j}^2$$

Also, Huang and Tauchen (2005) defined the Relative Jump, RJ_t as the contribution of jumps to the total variance, as follows:

$$RJ_t = \frac{RV_t - BV_t}{RV_t}$$

Through this definition, 100*RJ_t is equal to the percentage contribution of jumps, if any, to total price variance. All of these quantities, RV_t, BV_t, and RJ_t , are then totaled cumulatively throughout each trading day. For example, if 5-minute returns are used, there are 77 returns in each day; therefore all 77 values are used to generate a summation series for each statistic for each of the 1241 trading days in the sample.

Again, these values are calculated cumulatively over each trading day. These statistics are combined to calculate the z-statistic for each day, testing the null hypothesis that there were no jumps present during the day:

$$z_t = \frac{RJ_t}{\sqrt{(v_{bb} - v_{qq})\frac{1}{M}\max(1, \frac{TP_t}{BV_t^2})}},$$

where

$$v_{qq} \quad = \quad 2$$

and

$$v_{bb} \quad = \quad (\tfrac{\pi}{2})^2 + \pi - 3.$$

The z-statistic also utilizes the Tri-Power Quarticity, TP_t, which is defined as:

$$TP_t = M\mu_{4/3}^{-3}(\frac{M}{M-2})\sum_{j=3}^{M}|r_{t,j-2}|^{4/3}|r_{t,j-1}|^{4/3}|r_{t,j}|^{4/3},$$

where

$$\mu_a = \mathrm{E}(|Z|^a), \quad Z \sim N(0,1),\ a > 0.$$

Barndorff-Nielsen and Shephard also show that the Tri-Power Quarticity is jump-robust as an estimator of the integrated variance squared, as shown:

$$TP_t \rightarrow \int_{t-1}^{t} \sigma^4(s)ds.$$

Therefore, by using the ratio of the Tri-Power Quarticity to the Bipower Variation, the z-statistic simply follows the general form of any z-statistic, as defined below:

$$z = \frac{test - null}{\sqrt{\sigma_{test}^{\ 2}}}.$$

In this case, the z-statistic tests the null hypothesis that RJ_t is equal to 0. Therefore, the denominator of the z-statistic represents the square root of the variance of RJ_t. This version of the z-statistic is the recommended statistic presented in Huang and Tauchen's analysis of the theoretical statistic presented by Barndorff-Nielsen and Shephard. These *z* values are taken at the .1% significance level, in order to flag only large jumps. As stated before, these statistics only identify days on which there is evidence of at least one jump; they cannot show how many jumps are in each of these days.

B. The Lee-Mykland Return-to-Volatility Ratio Statistic

As stated before, the Lee-Mykland test relies on constructing a ratio of the current return and the local volatility. Using the price model, with *S(t)* being the price at time *t*,

$$d \log S(t) = \mu(t)dt + \sigma(t)dW(t) + Y(t)dJ(t)$$

where *Y(t)* is the jump size, and *dJ(t)* is a non-homogenous Poisson-type jump process. The first two terms are defined as a drift term, added to a Brownian motion term.

Therefore a t-statistic is proposed to test the null hypothesis that there is no jump at a given return, constructed as follows:

$$z_{t,j} = \frac{r_{t,\ j}}{\sigma_{t,j}},$$

where the z-statistic contains a moving average of the Bipower Variation to estimate the local volatility, constructed as

$$\sigma_{t,j} = \sqrt{\frac{\pi}{2} \times \frac{1}{K+1} \times \sum_{i=0}^{K} \; | r_{t,\, j-i} || r_{t,\, j-i-1} |} \, .$$

K in this formula represents the backward-looking window size, and $\frac{\pi}{2}$ is a constant used to normalize the statistic so that a z-table can be used. The subscript t is used to denote the day, while j shows the return on a given day. When the value of j is negative, the term then refers to a return on the previous day. Lee and Mykland recommend window sizes of 7, 16, 78, 110, 156, and 270 returns for sampling intervals of 1 week, 1 day, 1 hour, 30 minutes, 15 minutes, and 5 minutes, respectively.

Throughout this research, five minute returns were analyzed, so a window size of 270 was used. One important step in evaluating this statistic points out that a lower significance level must be used to account for Type I errors due to the much higher number of returns, a. Therefore, using a binomial distribution, setting $.999 = \Pr\,(k{=}0) = \binom{n}{k}(\alpha)^{n-k}(1-\alpha)^{k} = n(\alpha)^{n}$, where n is equal to the number of statistics in each sample, alpha is solved for, which becomes the adjusted significance level. Table 1 in Appendix B shows the suggested values for various sampling intervals. By using these values, it will be equivalent to using a .1% significance level on the daily level, as was used in the Barndorff-Nielsen and Shephard statistics. This table will be used later when applying this statistic to actual stock data.

C. Introduction of the Two-Stage Process

After examining the L-M test, which simply constructs a ratio of the current return to an estimate of the local volatility, it is proposed that it might be possible to make the statistic slightly more efficient. As Huang and Tauchen (2005) found, the Realized Variance is a more efficient estimator of local volatility than the Bipower Variation; however, this is only the case when a sample with no jumps is considered, since Realized Variance is not robust to jumps, as previously discussed. Therefore, it is reasoned intuitively that, if the Lee-Mykland statistic is able to flag enough jumps, the Realized Variance can possibly then be used to re-compute a more efficient estimate of local volatility after these potential jumps are removed.

The new statistic is constructed according to the following steps. First, the ratio of the current return to the local volatility estimate, computed using Bipower Variation, is calculated. This generates a series of returns flagged as jumps, which are then set to zero. Next, the local volatility throughout the sample is recalculated using Realized Variance with this new set of returns. Finally, another z-statistic is created, using a ratio of the original set of returns to the new estimate of local volatility. In this statistic, the estimate of local volatility will be defined as:

$$\sigma'_{t,j} = \sqrt{\frac{1}{K+1} \times \sum_{i=0}^{K} |r_{t,j-i}|^2} \ .$$

The application of this new test to Bristol-Myers stock data, as well as simulated data sets of returns, will be reviewed later in this paper.

III. Results of Application to Observed Stock Data

A. Analyzing High-Frequency Stock Data with the Barndorff-Nielsen Shephard Statistic

This section, as discussed in the introduction, will discuss the various analyses performed on actual BMY high-frequency stock. These analyses consisted of the Barndorff-Nielsen Shephard statistic, the Factivas news search on the flagged jump days, the Lee-Mykland statistic, and the new two-stage statistic.

First, the price of Bristol-Myers Squibb Co. stock is displayed. This graph is presented below (Figure 1a).

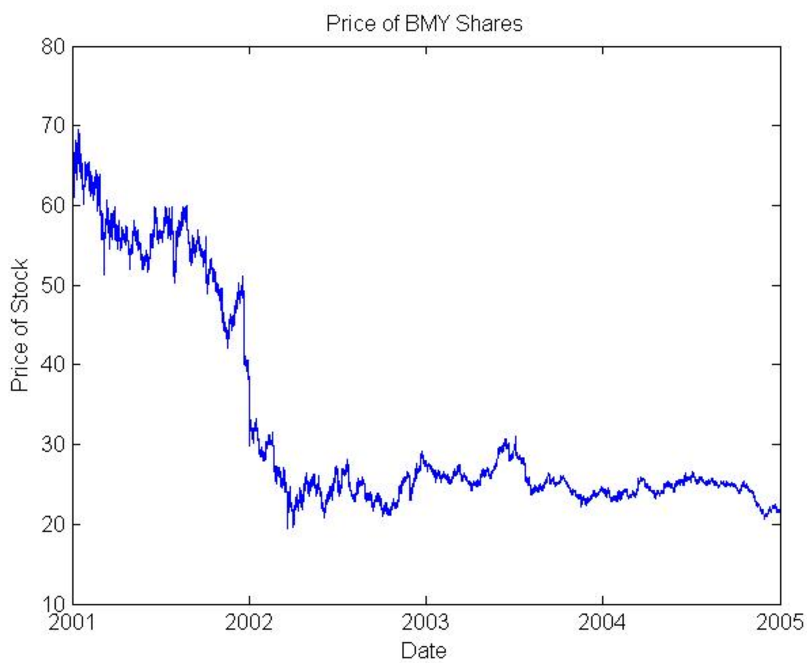

Figure 1a: Graph of BMY Stock Price from Jan. 2001 to Dec. 2005

Obviously, the enormous drop in the stock price stands out, and it is also noted that volatility seems to dramatically decrease during the last half of the sample. It is hypothesized that this large price change that occurred just before 2002 will be flagged as a jump. However, one cannot tell which price changes are overnight price moves. Since each day's returns are used to compute the Barndorff-Nielsen and Shephard statistic, all overnight returns are ignored.

Next, from this series of prices, the differences of the log-prices are taken to find the returns. Figure 1b shows the returns of BMY stock, shown in differences of the log-returns.

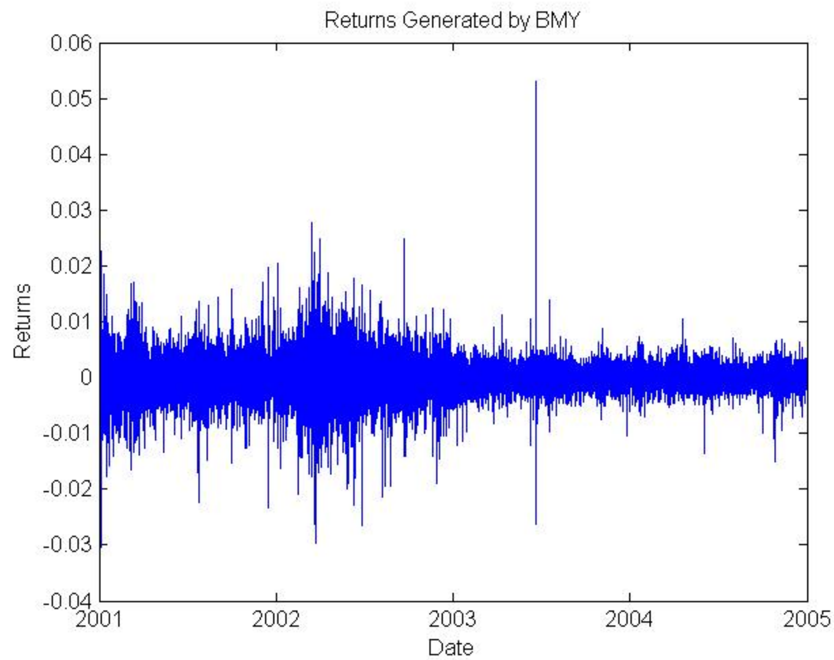

Figure 1b: Graph of BMY Stock Returns from Jan. 2001 to Dec. 2005

These returns will be used to compute the final Barndorff-Nielsen and Shephard statistic. Next, these returns are used to calculate the z-statistic, as defined earlier. Using a .1% significance level, 51 jump days are detected. The graph of the computed z-statistics is shown in Figure 1c. The horizontal line displays the cutoff value, corresponding to the .1% significance level.

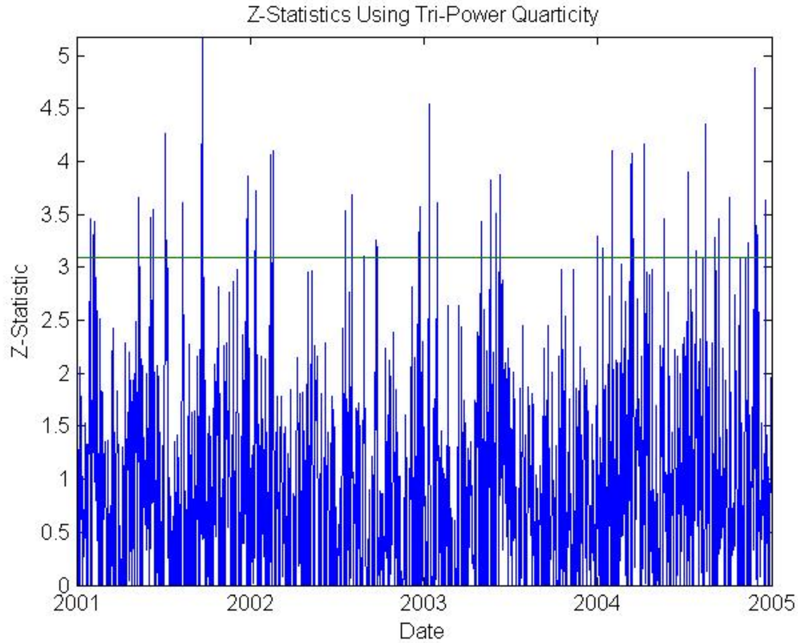

Figure 1c: Graph of Daily Z-Statistics Computed for BMY Stock from Jan. 2001 to Dec. 2005

B. Analyzing Flagged Jump Days Using Factiva

Next, after these fifty-one days were flagged as containing a jump, the Factiva news service was used to attempt to speculate what event might have caused the stock price to jump. By examining each of the days in such a way, the study could offer some clue as to what sorts of events might trigger a jump in the stock. After examining the results, there are five out of fifty-one flagged days that did not seem to be caused by a company-specific event occurring to Bristol-Myers, and there seem to be four major types of events on the remaining 46 flagged days: product liability or antitrust lawsuits against Bristol Myers, accounting or financial announcements released by Bristol-Myers, mergers or acquisitions, and product development news.

The least frequent of these, occurring on eight flagged days, pertains to product liability or antitrust lawsuits against Bristol-Myers Squibb. For example, the statistic flagged both June 2002, the company was sued for obstructing competitors, and February 14 2001, a potential investigation into Bristol-Myers by the FTC was announced.

The next category of announcements, falling on fifteen days flagged as jumps, concern accounting news or announcements concerning the financial health of the company, such as earnings announcements. For example, the statistic flagged March 10, 2003, when Bristol-Myers announced a restatement of financial statements. It also flagged August 9, 2005, 2005, when profit jumped 89.5%, exceeding projected earnings per share by $.14.

The third category of events, falling on seventeen flagged days that seemed to trigger stock price jumps, is news concerning changes in the structure of the company, such as mergers and acquisitions, or expanding drug distribution lines into new areas. For example, the statistic flags April 1, 2002 as a jump, the same day that Chinoin acquired a Hungarian drug unit from Bristol-Myers. The statistic also flags both January 5, 2005 and July 15, 2005, the days that Bristol-Myers' sale of its Excedrin line became announced and became finalized, respectively.

Finally, the most prevalent news event that seems to trigger stock price jumps, corresponding to twenty flagged days, concerns drug development news, whether good or bad. As one might expect, the future earnings estimates for a pharmaceutical company largely depend on such announcements, and the days flagged as containing jumps seemed to indicate this. For example, December 3, 2001, is flagged as a jump. This was the same day that IVAX, a competitor, gained approval to make a generic version of one of Bristol-Myers' most successful drugs, Glucophage. Also, the test declares July 23, 2003, when the Food and Drug Administration approved Bristol-Myers' Pravigard tablets, as a jump. In general, these sorts of events seemed to have a definite correlation with the days flagged by the z-statistic developed in Huang and Tauchen (2005).

Next, the research sought to visually confirm some of the results produced by the statistic. To do this, several of the days where drug development news seemed to cause a price jump were examined. Figures 2a-2e in Appendix A show five of these days- December 3, 2001, December 13, 2002, July 23, 2003, October 6, 2004, and November 21, 2005. Each of these graphs shows the returns throughout each trading day, in percent, and beneath each figure is a description of the drug announcement that occurred each day. These days all seem to have one or more returns that might visually indicate a large, sudden jump in the stock price. After this analysis was performed, the research moved to the Lee-Mykland statistic, which sought to be able to identify the specific return where the jump occurred, not merely the day where at least one such jump had been present.

C. Analyzing Actual Stock Data with the Lee-Mykland and Two-Stage Statistic

The data set is also analyzed using the Lee-Mykland statistic and the two-stage approach, as described previously. Using the .012% significance level found in Table 1 in Appendix B, appropriate for 5-minute returns, the normal Lee-Mykland statistic flags 663 returns as having a jump. When running the Realized Variance iteration, 804 returns are flagged in the second phase. Also, slightly different significance levels are tried during the first stage to adjust how many returns were set to zero before computing the jump-free variance. For example, when using a .1% significance level in the first stage (followed by the usual .012% significance level in the second stage), 1,339 jumps are flagged by the first stage, while the second stage detects 1,099 jumps. Also, while using a .05% significance level in the first stage, 1,057 jumps are flagged in the first stage, while the second stage flags 1,005 returns as jumps.

While it is very interesting to note that the statistic flags a great deal more jumps than Barndorff-Nielsen and Shephard's statistic, not many conclusion can be gathered from these results, since one does not know definitely how many jumps were in the sample, and whether or not adding the second stage increased the accuracy of the statistic. Therefore these statistics will be calculated on simulated data, where jump locations are known.

IV. Validation of the Two-Stage Statistic Under Controlled Conditions

A. Returns Modeled Using Constant Volatility

The first group of simulations is performed using a very simplistic way of modeling a stock's returns. This is done by creating one Brownian motion term and two jump terms. Each jump term is a random Poisson process multiplied by a random normally distributed variable. Each normally distributed random variable has equal means of opposite signs, so that the jumps have a combined mean of 0. The formula used is shown below:

$$dp(t) = \sigma * dW(t) + d\lambda_1(t)\alpha(t) + d\lambda_2(t)\beta(t),$$

where σ is distributed N(0,.00004), $d\lambda_1(t)$ and $d\lambda_2(t)$ are random Poisson integer processes with mean .01, and $\alpha(t)$ and $\beta(t)$ are distributed N(±.01, .001). These parameters are selected to match the standard deviation of returns in the actual BMY stock data, which is equal to .002 in the 5-minute data. The jumps flagged by the Lee-Mykland statistic are slightly smaller but not statistically different from 1% jumps, so $\alpha(t)$ and $\beta(t)$ were centered around ±1%. The means of the Poisson processes were selected to provide a given number of jumps in the simulated data. Using the above parameters, approximately 2% of the returns contain a jump, which is approximately the percentage of the returns that were flagged by the tests in the actual Bristol-Myers stock data.

This price model poses two major limitations. First, the volatility is constant throughout the sample, which is not the case in actual stock data; in reality, volatility changes over time. This limitation is changed in simulations that will be discussed later in this paper. The second limitation is brought about through the use of the random Poisson integer processes. In these processes, a series of integers is generated, having a given mean, in this case equal to .01. Due to the low magnitude of the mean, it is assumed that the vast majority of non-zero integers generated would be equal to 1; however, in the case where the integer generated is greater than one (i.e., x=2,3…), this is equivalent to saying that there is more than one jump in the given return period, and that they are of equal intensity. In reality, this might not be an accurate model of jumps as seen in stock returns; however, due to the low mean of the Poisson distributions, it is assumed that these effects would be negligible, and the model could be used.

Throughout this paper, the results of these simulations will be presented using confusion matrices. This method of analyzing the results presents the performance of test statistics such as these in a 2x2 matrix as shown:

		Test	
		(NJ)	(J)
Truth	(NJ)	1	2
	(J)	3	4

Table 1: Sample Confusion Matrix

This matrix, in which each entry will be a decimal between 0 and 1, represents a different aspect of the accuracy of the test statistic. On each side of the matrix, the titles (NJ) and (J) refer to "Non-Jump" and "Jump," respectively. Each column corresponds to how the test statistic classified each return, and each row designates what each return really is in the simulation. For example, entry 1 refers to the ratio of returns that did not contain a jump that were correctly labeled as non-jumps by the test statistic. Meanwhile, entry 2 refers to the ratio of returns that did not contain a jump but were declared as jumps by the statistic. Likewise, entry 3 shows the ratio of returns that did contain a jump that were not flagged by the statistic as having a jump, and entry 4 is equal to the ratio of returns that did have a jump that were correctly identified by the statistic. Entries 2 and 3 are referred to as the false positive and false negative error rates, respectively. Also, it should be noted that each row should sum to 1. Finally, for many sets of simulations conducted, two confusion matrices will be shown: the top will be the performance of the original Lee-Mykland statistic under the given conditions, while the bottom will show the results of the modified, two-stage statistic.

Overall, six different variations of the two-stage statistic are tested using this model to simulate data, as shown below in Tables 2a-2f. Each test is performed by averaging the results of tests consisting of 77,000 simulated returns each. First, the iterative Realized Variance method is utilized, setting each jump flagged by the first stage equal to the absolute average return over the sample when the jump-free estimated local volatility is calculated (Table 2a). Table 2b shows the results when each jump flagged by the first stage is set equal to zero.

In analyzing these results, two comparisons can be made. First, the effect of utilizing the two-stage approach can be examined. Also, one can compare the two methods of conducting the two-stage approach.

	(NJ)	(J)
(NJ)	1	0.0000
(J)	0.1102	0.8898
(NJ)	1	0.0001
(J)	0.072	0.9277

Table 2a

	(NJ)	(J)
(NJ)	1	0.0000
(J)	0.1102	0.8898
(NJ)	0.998	0.0002
(J)	0.0684	0.9316

Table 2b

Examining these results, there does seem to be an improvement from using the second stage approach, as the correct detection rate increases from almost 89% using the normal Lee-Mykland to approximately 93% in the second stage for each test; however, if the same significance level is used for both stages, as in Tables 2a-2b, the false positive rate is higher in the first stage than when using the original Lee-Mykland statistic. However, this increase is fairly small in these tests, rising from virtually 0% to .01% in the second stage. There does not, however, seem to be a large difference created by changing the method of removing jumps before calculating the second stage statistic.

Next, significance levels are manipulated, as Table 2c displays the results of lowering the significance level of the first stage to 1% (while maintaining a .01% significance level for the second stage). Table 2d displays the results obtained when the significance level of the first stage is set to .5%. It is proposed that this allows the first stage to flag more jumps and remove them from the data set, then let the second stage attempt to identify jumps with the original low significance level. Ideally, this would strike a balance in the trade-off between flagging more jumps correctly and introducing the higher false positive rate. The two matrices below display the results of the second stage after using a different significance level used in the first stage. The detection rates should be compared to the rates found in the second stage of Table 2a. Also, one should note the difference in the false positive rates between the first and second stage for each test.

	(NJ)	(J)
(NJ)	0.9969	0.0031
(J)	0.0624	0.9376
(NJ)	0.997	0.0003
(J)	0.0666	0.9334

Table 2c

	(NJ)	(J)
(NJ)	0.9786	0.0214
(J)	0.0523	0.9477
(NJ)	0.999	0.001
(J)	0.0627	0.9373

Table 2d

In both tables, the changed significance level enabled the second stage to identify more jumps than when both stages have a .01% significance level pushing the correct detection rate to well over 93%, while having a lower false positive rate than the first stage, which utilizes a higher significance level. For example, in Table 2c, the first stage flags about 93.8% of the jumps correctly, but has a .3% false positive rate. The second stage, using a lower significance after these jumps are flagged, as virtually the same correctly detection rate (93.3%), but has only a .03% false positive rate. This process of changing the significance level between stages will be examined further under stochastic volatility conditions.

Finally, Tables 2e-2f explore changing the exponents of the Bipower Variation term to one-half and three-halves, in an attempt to determine if weighting the previous and current return makes a significant difference. Because of the changed exponents, a different coefficient was used in order to normalize the z-statistic, so that, for the estimated local volatility in Table 2f and Table 2g, respectively,

$$\sigma'_{t,j} = \sqrt{1.4142 \times \frac{1}{K+1} \times \sum_{i=0}^{K} |r_{t,\,j-k}|^{.5} |r_{t,\,j-1-k}|^{1.5}}$$

and

$$\sigma'_{t,j} = \sqrt{1.4142 \times \frac{1}{K+1} \times \sum_{i=0}^{K} |r_{t,\,j-k}|^{1.5} |r_{t,\,j-1-k}|^{.5}}.$$

	(NJ)	(J)
(NJ)	1	0.0000
(J)	0.1624	0.8376
(NJ)	0.9998	0.0002
(J)	0.0661	0.9339

Table 2e

	(NJ)	(J)
(NJ)	1	0.0000
(J)	0.1605	0.8395
(NJ)	0.9999	0.0001
(J)	0.0713	0.9287

Table 2f

Interestingly, Tables 2e-2f, with the weighted-return approach, seem to indicate that the first stage is less efficient than the normal Lee-Mykland test, flagging just under 84% of the jumps correctly; however, the second stage seems to significantly raise the detection rate, making the overall accuracy approximately equivalent to the normal two-stage test statistic, flagging around 93% in both Table 2e and Table 2f. From these results, it was hypothesized that weighting the returns in such a way would not provide a significant improvement, and this method is not examined under stochastic volatility assumptions.

Overall, adding the second iteration does not seem to make an extremely large difference, although the second stage does flag slightly more jumps. It is expected that raising the significance level of the first stage can provide the largest increase in correctly identified jumps after the second stage. With this data, however, the increase seems to be fairly minimal, indicating that perhaps the higher efficiency of the Realized Variance is not enough to create a large increase in jump detection accuracy. This may be caused by the constant volatility and the very high accuracy of the first stage under constant volatility. As the next section will show, adding a time-varying volatility increased the effect of changing the significance level of the first stage.

B. Returns Modeled Using Stochastic Volatility

After creating this very simple model with constant volatility, a more complex model was introduced. In an attempt to introduce stochastic volatility to provide a more realistic approximation of actual stock returns, the following model for the simulated returns is modified to include stochastic volatility:

$$dp(t) = \sigma_t * dW(t) + d\lambda(t)\alpha(t) + d\lambda 2(t)\beta(t)$$

A fairly basic model of volatility is used to simulate σ_t, adhering to the following model:

$$d\sigma_t(t) = k * (\sigma_{t\,avg} - \sigma_{t-1}) + \sigma_v dW_2(t) \, .$$

This volatility model consists of two terms. The first, $k * (\sigma_{t\,avg} - \sigma_{t-1})$, is a mean-reverting term. The *k* term is equivalent to the persistence of a given "jump" in volatility, represented in the percentage of the jump that decays after one return. For example, if k=.01, the change in volatility will revert 1% closer to its mean volatility each return. Therefore, the smaller k is, the more persistent any volatility change will be. The second term, $\sigma_v dW(t)$, represents the standard deviation of the standard deviation. Also, if at any time this process caused σ_t to become negative, the absolute value was taken. By combining these two terms, the second term will create shocks that will be persistent yet mean-reverting, similar to actual financial markets. The graph below (Figure 3) is a sample graph of the standard deviation of the returns in a randomly generated sample of returns. As one can see, there is a large amount of movement and stochastic volatility present.

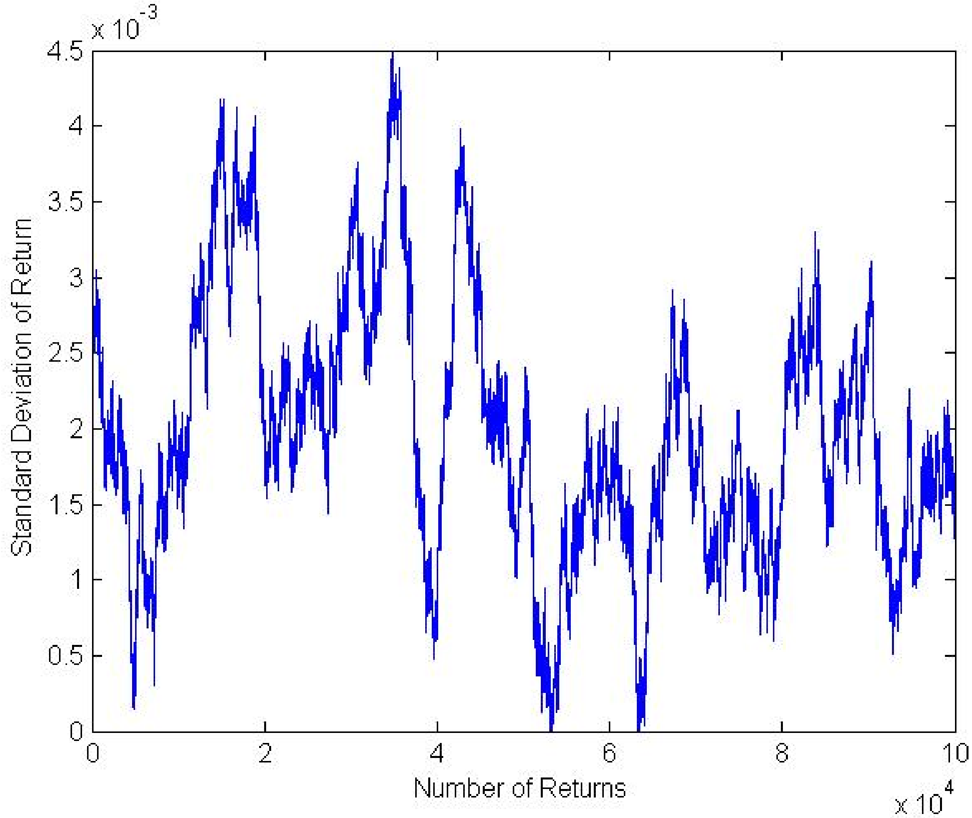

Figure 3: Changing Standard Deviation of Returns Throughout a Simulated Sample

In addition to evaluating the effect of adding the second stage of the statistic, adding this type of return modeling allowed for the analysis of the performance of the original Lee-Mykland statistic under certain conditions. For example, while attempting to calibrate the values of σ_v to use in a model, an interesting relationship arose between a given σ_v value and the percentage of jumps that both the Lee-Mykland and two-stage statistic identified: the accuracy of each of the tests is extremely dependent on the value of σ_v *for a given k value*. This fact is illustrated in the figures below. Figure 4 shows the required σ_v values in order for the true positive rate to stabilize.

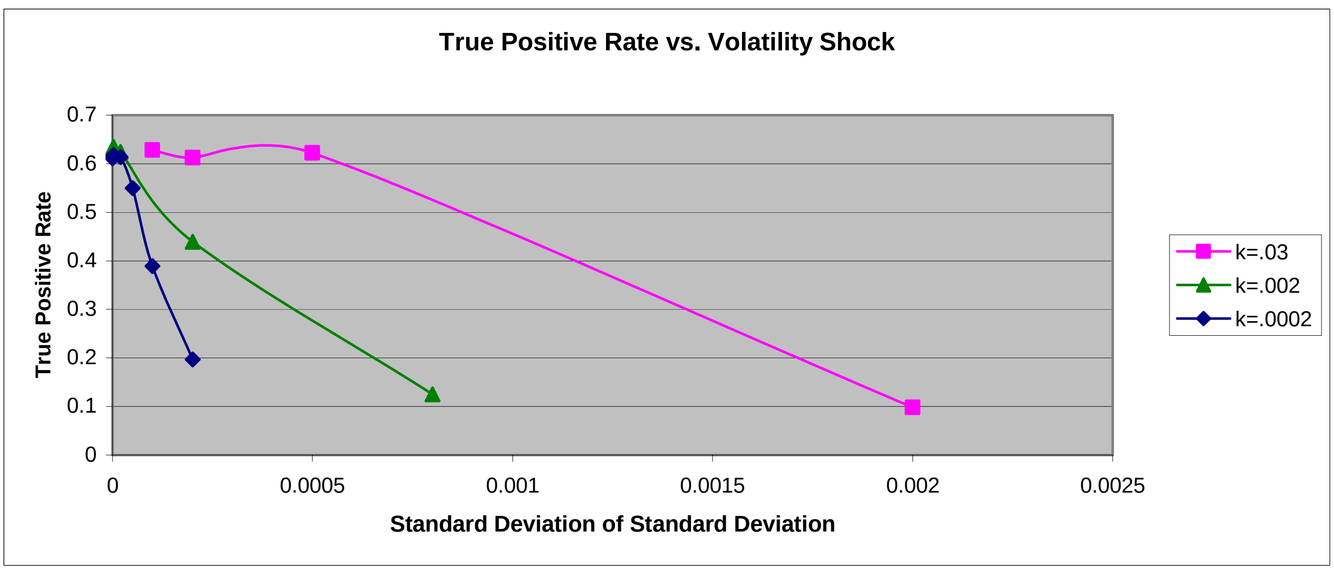

Figure 4: Graph of True Positive Rate vs. σ_v, Multiple k Values

As one can see, the difference in required σ_v values in order for the statistics to flag approximately 60% of the jumps correctly is vastly different for different persistence values. If volatility shocks have very high persistence, the statistics require a very low magnitude of σ_v in order to pick up a high percentage of the jumps. However, once a given σ_v is low enough to reach this, the accuracy of the statistics seems to plateau. This fact is an interesting observation of the conditions under which the statistics seem to be viable.

In order to calibrate the σ_v value, the kurtosis of the simulated returns is examined. First, the persistence level is calibrated so that volatility shocks would have a half-life of between 40-50 days. This criterion is chosen because recent research has placed the half-life of such shocks as being very large, as discussed in papers such as Bollerslev and Mikkelsen (1996). Once the persistence is set, a value of σ_v is selected that would result in a kurtosis of between 10 and 15, a range of values that is typical of the majority of stocks. For this analysis, it should also be noted that this σ_v value was lower than the plateau point for the given k value, approximately represented by the $k=.0002$ data set in Figure 4.

First, the statistics are examined using different significance levels (but keeping the same significance level for each stage). Table 3a shows the confusion matrices with .01% significance levels, while Tables 3b-3c display the effects of using .1% and 1% significance levels for each stage, respectively. As one can see, there is a large increase in the positive detection rate as the significance level increases; however, this is accompanied by large increases in the false positive rate. Also, the increase in correct detection rates between the original L-M statistic and the two-stage statistic stays approximately between a 1-2%. For example, the correct detection rate rises from 60.6% to 61.7% in Table 3a. Since this difference is so small, there does not appear to be a large advantage in incorporating the second stage. One should pay particular attention to Table 3a, as these are the results using the standard, calibrated parameter values and significance levels that were previously discussed. These results will be used to compare the results each time a parameter is changed. Under these base conditions, the second stage correctly identified about 61.6% of the jumps, an increase of about 1% over the original Lee-Mykland statistic, while introducing a false positive rate of .04%.

	(NJ)	(J)
(NJ)	0.9997	.0003
(J)	0.3938	0.6062
(NJ)	0.9996	0.0004
(J)	0.3834	0.6166

Table 3a

	(NJ)	(J)
(NJ)	1	.00022
(J)	0.309	0.6915
(NJ)	0.997	0.0027
(J)	0.297	0.7026

Table 3b

	(NJ)	(J)
(NJ)	0.9834	0.0166
(J)	0.2495	0.7505
(NJ)	0.9725	0.0275
(J)	0.2225	0.7774

Table 3c

Next, the significance levels only for the first stage are changed in an attempt to minimize the trade-off between raising the positive detection rate while increasing the false negative rate. The intuitive reasoning behind this was such: if the significance level for the first stage is raised, it will flag more returns as jumps. If more returns are set to 0, the second stage, still utilizing a significance level of approximately .01% will be more likely to flag jumps that were just below the cutoff threshold previously. Tables 4a-4c display the results when the first stage uses .1%, 1%, and 5% significance levels, respectively. These seem to provide a viable way to accurately declare more jumps while maintaining a fairly low false negative rate, particularly Table 4b, which uses a 1% significance level in the first stage. These data show an increase of nearly 7% in the correct detection rate from the original rate of 61.7% to 68%, while raising the false negative rate only .05% to .09%.

	(NJ)	(J)
(NJ)	0.9979	0.0021
(J)	0.3339	0.6661
(NJ)	0.9995	0.0005
(J)	0.3705	0.6295

Table 4a

	(NJ)	(J)
(NJ)	.9833	.0167
(J)	0.2269	0.7731
(NJ)	0.9991	0.0009
(J)	0.3197	0.6803

Table 4b

	(NJ)	(J)
(NJ)	0.917	0.0830
(J)	0.1815	0.8185
(NJ)	0.9968	0.0032
(J)	0.3159	0.6841

Table 4c

Next, the average jump size in the sample is changed. The original parameters, as in the constant volatility case, set the absolute average jump equal to 1%. While this is a very high jump when compared with some of the data, it is possible that in portions of returns with high volatility, these jumps may be washed out and might go undetected by the statistics. Therefore, jump sizes are adjusted to examine the statistics' performances under these different conditions. Tables 5a-5c show the results when the absolute average jump sizes are .5%, 1.5%, and 2%, respectively. As expected, as jump size increases, the statistics are able to pick up the jumps much more accurately. These results are expected, but it is interesting to see how poorly the statistics perform with .5% jumps being added (a 48.06% correct detection rate), and how well they perform with 2% jumps added (an 87.52% correct detection rate). Also, similar to Tables 3a-3c, the second stage only increases the true jump detection rate by about 1%, indicating a negligible effect of adding the second stage.

	(NJ)	(J)
(NJ)	0.9996	0.0004
(J)	0.5265	0.4735
(NJ)	0.9997	0.0003
(J)	0.5194	0.4806

Table 5a

	(NJ)	(J)
(NJ)	0.9997	0.0003
(J)	0.2567	0.7433
(NJ)	0.9997	0.0003
(J)	0.2457	0.7543

Table 5b

	(NJ)	(J)
(NJ)	0.9997	0.0003
(J)	0.1352	0.8648
(NJ)	0.9996	0.0004
(J)	0.1248	0.8752

Table 5c

Finally, the jump frequency was adjusted. In the earlier stochastic volatility models, jumps arrived, on average, 1 in every 100 returns. Since the data is being calibrated to five-minute returns, this means about one jump every day and a half. Tables 6a-6c adjust this frequency for jumps to arrive every 50, 25, and 10 returns, respectively. There are two very important observations that can be gathered from this data. First, both statistics pick up fewer jumps correctly on the whole as jumps begin arriving more frequently. This can be seen clearly, as the first stage only correctly identifies 40% of the jumps in Table 6c; however, as jumps arrive more frequently, the second stage begins to have a larger effect. For example, in Table 6c, the second stage increases the correct detection rate over 7% to 47.24%, while maintaining virtually a 0% false positive rate. As the first stage starts to have more trouble flagging jumps correctly as they arrive more frequently, the second stage can identify jumps much more accurately once some of them are removed by the L-M statistic. It appears that this is the second case in which the second stage provides an advantage to Lee and Mykland's statistic.

	(NJ)	(J)
(NJ)	0.9998	0.0002
(J)	0.3823	0.6177
(NJ)	0.9997	0.0003
(J)	0.3576	0.6424

Table 6a

	(NJ)	(J)
(NJ)	1.0000	0.0000
(J)	0.4499	0.5501
(NJ)	0.9999	0.0001
(J)	0.4001	0.5999

Table 6b

	(NJ)	(J)
(NJ)	1.0000	0.0000
(J)	0.5981	0.4019
(NJ)	1.0000	0.0000
(J)	0.5276	0.4724

Table 6c

IV. Discussion and Conclusions

Overall, this research has led to many interesting conclusions. First, the research on high-frequency stock data of Bristol-Myers Squibb Co., utilizing the Barndorff-Nielsen and Shephard statistics, combined with the Factiva News Service, attempts to examine connections between what these statistics indicate and real world events. There seems to be very strong connection between the two, with most of the days flagged as containing at least one jump having a corresponding news event that could have definitely triggered large moves in the stock price. Also, over 90% of these events are company-specific events, rather than events that seem to affect the market as a whole. While a larger study of many other companies' stock jumps is necessary the types of events that seem to cause these jumps (product liability or antitrust lawsuits, financial announcements, mergers or acquisitions, and product development news) might be very similar across many other stocks.

Next, the research shifts focus to Lee and Mykland's statistic, which seeks to identify the specific returns where jumps occur. When the two-stage variation of the L-M statistic is first proposed, it seems that the second stage might be able to greatly improve the accuracy in detecting jumps. While creating simulated returns with both constant and stochastic volatility, in most situations, there is not a large improvement when the second stage is used. This seems to indicate that the L-M statistic, using Bipower Variation to estimate local volatility, is a fairly efficient estimator, and that the Realized Variance used in the second stage most likely does not increase the accuracy significantly under most conditions. While it has been proven to be a more efficient estimator of local volatility than Bipower Variation in other papers, its higher efficiency does not result in dramatic accuracy gains in these simulations under most conditions.

However, the second stage seems to present possible benefits in two situations. First, when the significance level used in the first stage is raised while the significance level in the second stage is held constant (Tables 4a-4c), the false positive rate can be held at a low level by the low significance level in the second stage, while correctly identifying more jumps because of the higher significance level in the first stage. This method could hold significant promise for increasing the accuracy of the Lee-Mykland jump detection test. While it is not within the scope of this paper to speculate on such things, this iteration process could possibly be used on many other statistics which utilize the Bipower Variation to attempt to flag jumps in prices series.

In addition, the two-stage statistic seems to have a clear advantage over the original Lee-Mykland test when jumps in the returns are extremely concentrated, as seen in Tables 6a-6c. In these periods, the two-stage process allows many more jumps to be filtered out in the trailing average local volatility estimate for a given return. This then allows the second stage to more accurately identify jumps. However, price series with so many large moves in a stock are extremely rare. Therefore, this research concludes that, while there are some instances when adding the second stage of the statistic is beneficial, under most circumstances, the original Lee-Mykland statistic is fairly accurate, and the second stage does not greatly increase the test's accuracy unless significance levels are changed. However, if the first stage is able to flag a large number of jumps, either by raising the significance level used or simply by examining a set of stock prices where there are many jumps present, significant increases in the correct detection might be possible without greatly increasing the false positive rate.

Appendix A- Figures

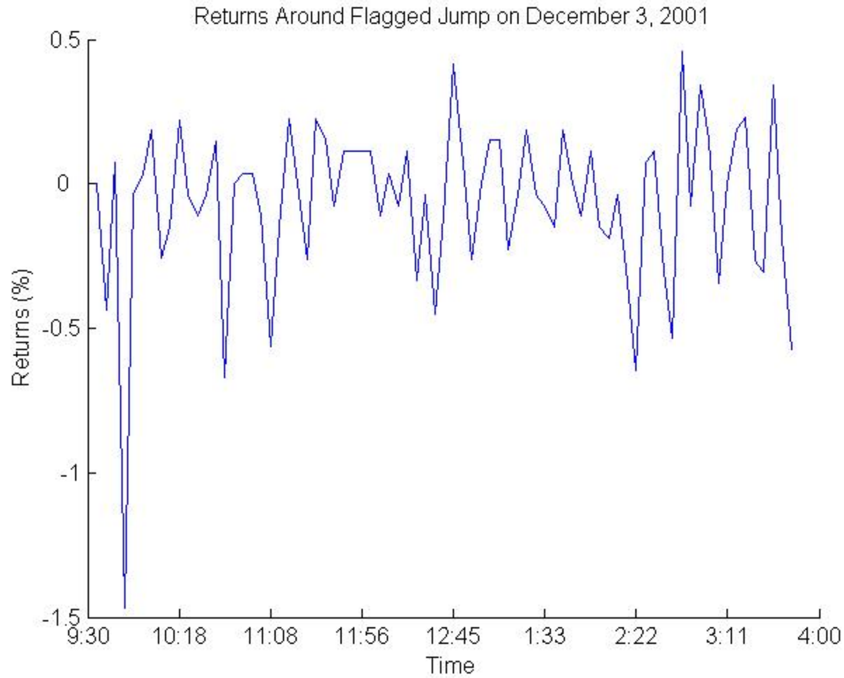

Figure 2a: Graph of returns on December 3, 2001, the day that IVAX, a BMS competitor gained approval for a generic rival to Bistol Myers' drug Glucophage. The jump certainly seems to stick out, occurring at approximately 9:50 am.

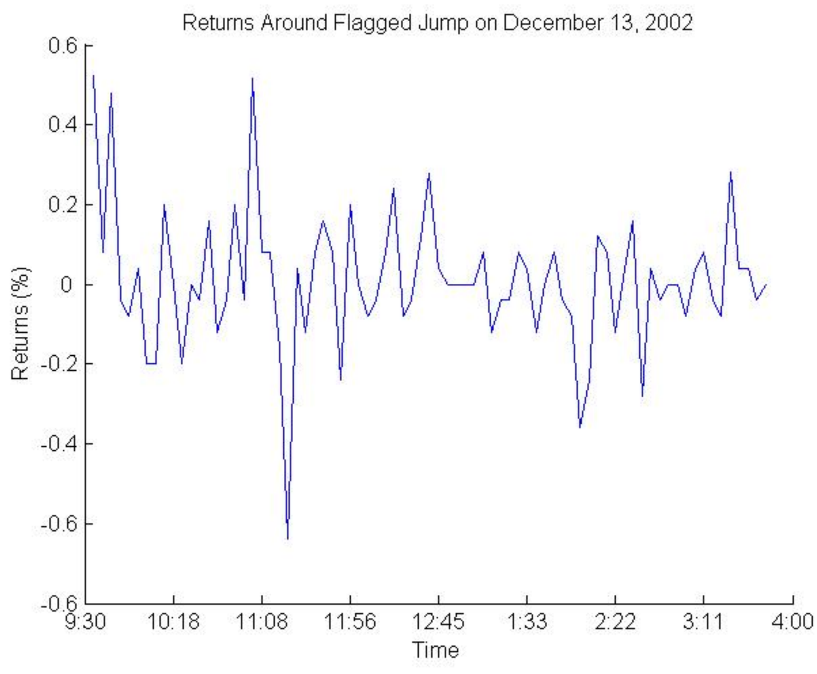

Figure 2b: Graph of returns on December 13, 2002, the day that the FDA approved Metaglip, a BMS drug. The jump in BMY stock seems to occur with either the large positive or negative move occurring from around 10:45 to 11:15 am.

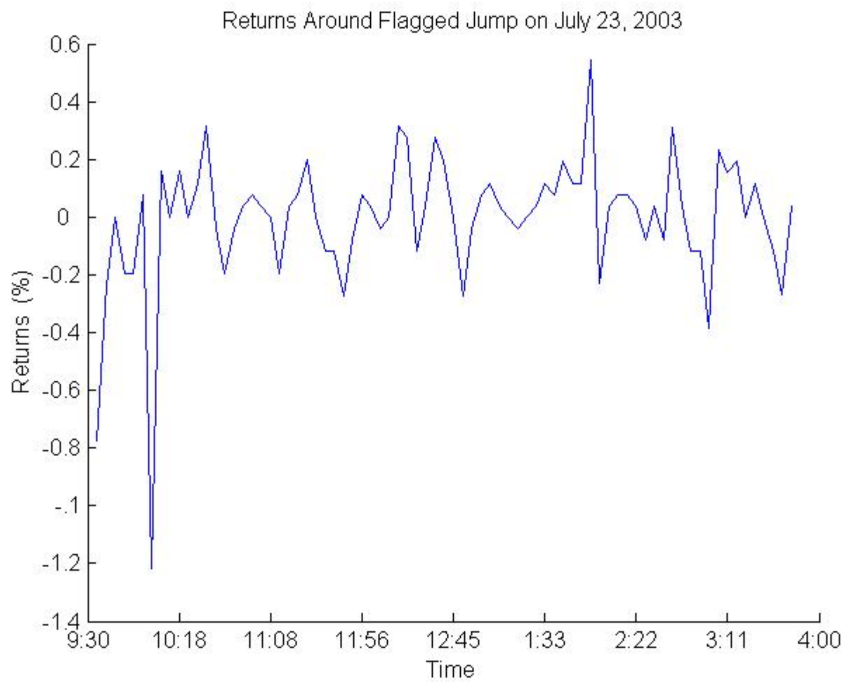

Figure 2c: Graph of returns on July 23, 2003, the day that the FDA approved Pravigard tablets, a BMS drug. Despite a huge early morning loss, due to steady growth, along with a large jump at approximately 2:00 pm, BMY stock was nearly even on the day.

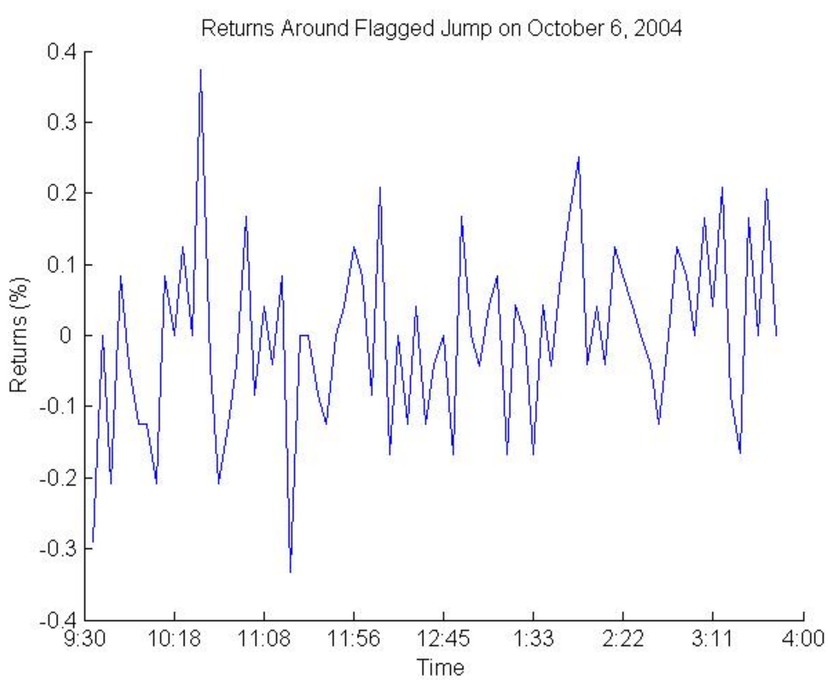

Figure 2d: Graph of returns on October 6, 2004, the date of a positive drug test announcement for BMS. The jump could be the large move at approximately 10:30 am.

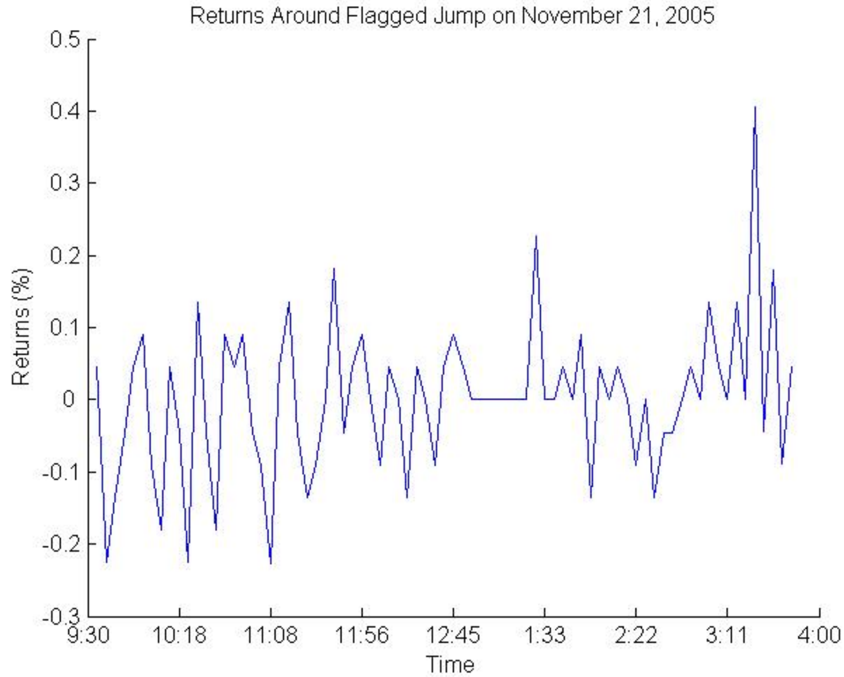

Figure 2e: Returns on November 21, 2005, the date of a positive test announcement of Bristol-Myers' drug Baraclue. It seems that this jump may be shown by the large move at approximately 3:20 pm.

Appendix B- Tables

Flagged Jumps at Recommended Window Sizes for SPY Data*

Sampling Frequency	*Window Size*	*Significance Level*
5 minutes	270	0.99988
7 minutes	250	0.999836
11 minutes	200	0.999753
17.5 minutes	150	0.999624
27.5 minutes	110	0.999435
38.5 minutes	90	0.999236
55 minutes	78	0.998948
77 minutes	75	0.9985758

Table 1: Significance Levels for Lee-Mykland Based on Sampling Interval

Works Cited

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. (2002). Parametric and Nonparametric Volatility Measurement. *Handbook of Financial Econometrics*. Amsterdam, North Holland: (eds.) Yacine Ait-Sahalia and Lars Peter Hansen (eds.).

Black, Fischer and Myron Scholes. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy,* 3, 637-654.

Barndorff-Nielsen, Ole E., Svend Erik Graversen, Jean Jacod, Mark Podolskij, and Neil Shep-hard. (2005). A Central Limit Theorem for Realised Power and Bipower Variations of Continuous Semimartingales. Working Paper, Nuffeld College, Oxford University; Yu Kabanov and Robert Liptser (eds.)..

Barndorff-Nielsen, Ole E., Svend Erik Graversen, Jean Jacod, and Neil Shephard. (2005). Limit Theorems for Bipower Variation in Financial Econometrics. Working Paper, Nuffeld College, Oxford University.

Barndorff-Nielsen, O.E., and N. Shephard. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2, 1-48.

Bollerslev, T. and Mikkelsen, H. (1996). Modeling and Pricing Long-Memory in Stock Market Volatility. *Journal of Econometrics*, Vol.73, No.1, pp.151-184.

Dacorogna, M., R. Gencay, U. Muller, R. Olsen, and O. Pictet. (2001). *An Introduction to High Frequency Finance*. San Diego: Academic Press.

Huang, X. and G. Tauchen. (2005). The Relative Contribution of Jumps to Total Price Variance, *Journal of Financial Econometrics* 3, 456-499.

Law, Tzuo Hann. (2007). The Elusiveness of Systematic Jumps. Undergraduate Honors Thesis, Duke University.

Lee, S. and A. Mykland. (2006). Jumps in Real-Time Financial Markets: A New Nonparametric Test and Jump Dynamics. *University of Chicago Technical Report Series*, No. 566.