Spurious Jump Detection and Intraday Changes in Volatility

Matthew Rognlie

Professor George Tauchen, Faculty Advisor

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

Duke University

Durham, North Carolina

2010

The Duke Community Standard was upheld in the completion of this thesis.

Abstract

We investigate the properties of several nonparametric tests for jumps in financial markets. We derive a theoretical property of these tests not observed in any of the previous literature: when they are applied to finitely sampled data, they are generally biased toward finding too many jumps. This results from bias in finite-sample estimation of several important test components. The severity of the bias corresponds to the magnitude of change in volatility over the course of a day. We use data on an intraday volatility pattern in several US equities, which results in quantitatively significant changes in the level of volatility during the day, to undertake Monte Carlo simulations of a price process without jumps. Applying several jump tests to the simulated data, we detect one-half to two-thirds as many jumps as in the observed data, suggesting that many jumps currently detected in empirical applications of these tests are spurious. We also present several possible modifications to jump tests that limit the effect of intraday patterns in volatility, all of which produce dramatically lower estimates of the frequency and importance of jumps.[1]

[1] I would like to thank Professors George Tauchen and Tim Bollerslev for all their help, advice, and encouragement. I would also like to thank my peers in the Honors Workshop—Pongpitch Amatyakul, Sam Lim, Abhinay Sawant, and Derek Song—for perceptive commentary throughout the semester. Finally, I am grateful to Andrey Fradkin and Peng Shi for helpful conversations.

1 Introduction

Recent literature suggests that discontinuities, or "jumps," are an essential part of financial price changes. For instance, Andersen, Benzoni, and Lund (2002) find that any reasonably descriptive continuous-time model for equity index returns must allow for discrete jumps. Maheu and McCurdy (2003) present evidence that a model incorporating jumps can improve forecasts of volatility, and Drost, Nijman and Werker (1998) develop a statistical test for the hypothesis that a series is generated by a continuous diffusion process, which strongly indicates the presence of jumps in dollar exchange rates when applied to data. Though jumps may arise from a variety of causes, ever since Merton (1976) a common explanation has been the sudden availability of new information. In principle, an efficient market will incorporate unanticipated news instantaneously, leading to a discontinuous change in the price of affected assets.

Jumps have numerous implications in pricing and risk management. Zhang, Zhou, and Zhu (2009) show that the jump risk of firms, as estimated from high-frequency equity prices, is a major predictive component of the premium on firms' credit default swaps. Andersen, Bollerslev, and Diebold (2007) observe that many common approaches to estimating volatility rely on the assumption of a continuous sample path, which is clearly violated in practice. In general, jumps complicate the derivatives pricing problem (see Kou, 2002), and make general equilbrium pricing models less tractable.

Given the theoretical relevance of jumps, it is important to be able to detect them in data. Several authors have proposed nonparametric statistical tests that determine whether a particular time interval contains a jump, or whether an individual price movement is likely to reflect a jump. Barndorff-Nielsen and Shephard (2004, 2006) distinguish between two measures of integrated variance, one jump-robust and one not, that together offer a way to test whether a sample contains jumps. Jiang and Oomen (2008) exploit the higher-order sample moments of returns to identify periods that contain jumps, and Ait-Sahalia and Jacod (2007) examine the difference between higher-order moments computed at two different sampling frequencies. Attempting to identify whether individual price changes are jumps, Lee and Mykland (2008) compare the magnitude of each change with a sliding-window measure of local volatility.

Although these tests are all designed to distinguish jumps from the diffusive component of volatility, some recent work suggests that they produce incoherent results. Schwert (2009) finds that tests proposed by different authors identify different days that contain jumps. Even more alarming, he also finds that tests are not even consistent with themselves, detecting different jumps when the sampling frequency is adjusted. For instance, one measure derived from Barndorff-Nielsen and Shephard detects jumps on 6.9% of days at 10-minute sampling and 6.4% of days at 15-minute sampling, but only 1.21% of days are flagged at both the 10-minute and 15-minute frequencies. Most jumps detected at one frequency, therefore, are not identified at another, suggesting that the results do not consistently reflect jumps in the actual data.

We provide a simple explanation that accounts for many of these contradictory results: dramatic intraday changes in volatility, combined with a coarse sampling frequency that distorts the asymptotic properties of our estimators, cause the jump statistic to detect jumps even in a completely diffusive price sequence. Half or more of the jumps detected in actual data appear to be artifacts of this behavior, and existing estimates of the empirical significance of jumps may be dramatically overstated.

Intraday patterns in volatility have long been observed in the literature. Wood, McInish, and Ord (1985) document U-shaped patterns in both the volatility and volume of equities, and their results are confirmed by Lockwood and Linn (1990). Generally, volatility is highest in the early morning at the beginning of the trading day. It declines until a minimum is reached in the early afternoon, at which point volatility begins climbing until the close. The average volatility at the peak is often twice or more the average volatility at the minimum.

Moreover, difficulties in jump detection arising from intraday volatility patterns are not new. Van Tassel (2008) shows that the test proposed by Lee and Mykland (2008) to detect whether specific price changes are jumps produces inconsistent results throughout the day. Specifically, the test exaggerates the number of statistically significant jumps in the early morning, when volatility is generally highest, while underreporting jumps in the middle of the day, when volatility is lower. This inconsistency arises from the test's use of a sliding window of returns to measure local volatility, which is then used to standardize returns and identify outliers that arise from jumps. To achieve correct results, returns in the morning should be standardized by a higher volatility measure than returns in the afternoon, but the Lee and Mykland test makes no allowance for these differences. Van Tassel finds that, as a result, almost half of jumps identified by the Lee and Mykland test occur during a small interval in the early morning. These results suggest that the effect of intraday volatility is an important consideration when evaluating nonparametric jump tests.

The effect examined in this paper, however, is both more general and of different origins. It arises from the fact that nonparametric jump tests are only valid asymptotically. For instance, the Barndorff-Nielsen and Shephard tests use an approximation to the integrated variance of a price process. To the extent that data is available only at non-infinitesimal intervals, practical applications of the Barndorff-Nielsen and Shephard tests—and many similar nonparametric tests—will suffer from *discretization error*. This is well known—see, for instance, discussion in Andersen and Benzoni (2008). There has been no discussion in the literature, however, of the fact that when the tests are run with finitely spaced data, many of their components are also systematically underestimated. In other words, finite-sample estimates of these components are not merely imprecise but also biased, which leads to an overestimation of the jump statistics themselves. This raises the possibility that practical applications of these tests misrepresent the importance of jumps.

We show that this bias has its origins in a simple mathematical inequality, and that the amount of bias depends on how quickly volatility changes throughout the sample period. Since the intraday volatility pattern is responsible for large changes in volatility over the course of a day—the sample period most often used for these tests—it is important to understand whether the resulting bias is quantitatively significant in the application of jump tests. We answer this question using simple Monte Carlo simulations, where price series are simulated under a jump-free stochastic model that incorporates the intraday volatility pattern, and another jump-free stochastic model where volatility is assumed to be constant. Applying jump tests to both sets of simulated data, we find that many more jumps are detected in the first set, even though both sets are generated by simulating a process without jumps. In fact, depending on the sampling interval and the type of jump test, the number of jumps detected in the first set of simulated data can be a sizable fraction of the number detected in observed data. At 15 minutes, an interval chosen to limit microstructure noise, it is one-half to two-thirds of the number detected in observed data, suggesting that many of the observed jumps may in fact be statistical artifacts.

This extends earlier work by Huang and Tauchen (2005), who use Monte Carlo simulation to investigate the finite sample properties of jump tests derived from Barndorff-Nielsen and Shephard (2006). Using two different volatility models to generate simulated price series, they conclude that the empirical size of the tests exceeds the nominal size, especially for the more complicated two-factor volatility model. This disparity, however, is not very large in practical terms, and Huang and Tauchen conclude that the tests perform impressively on simulated data. The results in this paper differ because of the introduction of the intraday volatility pattern. As the theoretical discussion will show, changes in volatility over the sample period are directly responsible for biased test results. Since the intraday volatility pattern produces swings in volatility much larger than those produced by typical calibrations of stochastic volatility models, our simulations capture a source of bias absent in earlier models.

The remainder of the paper of the paper proceeds as follows. First, in Section 2, it discusses standard stochastic models of stock price evolution. In Sections 3.1 and 3.2, it describes how Barndorff-Nielsen Shephard (BNS) and Jiang Oomen jump statistics are calculated. Next, in Section 3.3 it calls upon existing literature to establish the necessity of using staggered returns to compensate for the effects of microstructure noise. Continuing the theoretical discussion of jump test statistics, it shows in Section 4 that many of the components of these statistics are biased in finite samples with large changes in volatility, possibly leading to overdetection of jumps. Section 5 discusses the strong intraday pattern in volatility that is evident in the data, and relates this finding to the earlier discussion about the effects of changing volatility on our test statistics. Section 6 outlines some simulations and empirical work that will test the susceptibility of the jump statistics to intraday swings in volatility. Section 7 describes the details of the high-frequency pricing data that we will use, and Section 8 provides the results. Section 9 suggests several modifications to limit the damaging effects of the intraday volatility pattern, and provides empirical results from these modifications that substantiate our earlier findings. Finally, Section 10 draws some general conclusions from this work.

2 Stochastic Models of Returns

Consider a standard stochastic model of stock price evolution, given by a stochastic differential equation for log-prices $p(t)$:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \tag{1}$$

Here, $\mu(t)dt$ represents the time-varying drift component of prices, while $\sigma(t)dW(t)$ represents the time-varying volatility component, where $W(t)$ is standard Brownian motion and $\sigma(t)$ is the volatility level. Brownian motion $W(t)$ can be viewed as the limit from summing independently and identically distributed log-returns over infinitesimally small periods. The volatility level $\sigma(t)$ scales these returns to account for the width of the distribution of log-returns at time t.

This model of stock price evolution, however, produces continuous price sequences with probability one, which is inconsistent with empirically observed discontinuities in prices.

To incorporate these discontinuities, or "jumps," into our model, we add an additional term:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t) \tag{2}$$

where $q(t)$ is a counting process that increments by one with each jump and $\kappa(t)$ gives the magnitude of each jump. $dq(t)$ is hence the number of jumps in the infinitesimal interval dt.

While examining high-frequency data over relatively short time intervals, the drift process is generally insignificant enough to ignore. Unfortunately, it is often difficult to separate contributions to log-returns into the other two components: the jump process and continuous variation. We discuss two families of statistical tests that attempt to isolate significant jumps.

3 Jump Tests

3.1 Barndorff-Nielsen Shephard Tests

Barndorff-Nielsen and Shephard (2004) propose a test that compares two measures of variance to determine whether there is a statistically significant jump component during the sample period. The first measure, *Realized Variance*, converges as the sample frequency approaches infinity to the integrated variance plus a jump component, while the second measure, *Bipower Variation*[2], converges to the integrated variance alone. Formally:

$$RV = \sum_{i=2}^{n} r_i^2 \longrightarrow \int_0^T \sigma^2(s)ds + \sum_{i=1}^{n} \kappa^2(t_i)dq(t_i) \tag{3}$$

$$BV = \mu_1^{-2}\left(\frac{n-1}{n-2}\right)\sum_{i=3}^{n} |r_i||r_{i-1}| \longrightarrow \int_0^T \sigma^2(s)ds \tag{4}$$

where $r_i = p(t_i) - p(t_{i-1})$ is the geometric return from time t_{i-1} to time t_i, and $\mu_a = E(|Z|^a)$ for $Z \sim N(0,1)$. The times $t_1, \ldots, t_n$ are generally chosen to be equally spaced over the time interval $[0,T]$.

Clearly, the asymptotic difference between RV and BV will the jump component of variation. To test the significance of the detected jump component, however, we need to find the conditional standard deviation, which requires the *Integrated Quarticity* $\int_0^T \sigma^4(s)ds$. Anderson, Bollerslev, and Diebold (2007) propose using the realized *Tripower Quarticity* statistic to estimated integrated quarticity, while Barndorff-Nielsen and Shephard (2004) suggest the *Quadpower Quarticity* estimator:

$$TP = n\mu_{4/3}^{-3}\frac{n-1}{n-3}\sum_{i=4}^{n} |r_i|^{4/3}|r_{i-1}|^{4/3}|r_{i-2}|^{4/3} \longrightarrow \int_0^T \sigma^4(s)ds \tag{5}$$

$$QP = n\mu_1^{-4}\frac{n-1}{n-4}\sum_{i=5}^{n} |r_i||r_{i-1}||r_{i-2}|r_{i-3}| \longrightarrow \int_0^T \sigma^4(s)ds \tag{6}$$

Combining our estimates of integrated variance and integrated quarticity, we can make several possible test statistics. According to simulations by Huang and Tauchen (2005),

[2]The formula we give here is slightly different from the typical formula for Bipower Variation, with n and $n-1$ having been replaced by $n-1$ and $n-2$, respectively. The two expressions are asymptotically equivalent, but we choose the latter because it is more natural in the finite-sample context and is unbiased in the important limiting case of constant volatility. We make similar modifications to several other statistics.

however, the following *max-adjusted* statistics (which are asymptotically standard normal) perform best:

$$Z_{\mathrm{RJ,TP}} = \frac{RJ}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right)\left(\frac{1}{n}\right)\max\left(1, \frac{TP}{BV^2}\right)}} \tag{7}$$

$$Z_{\mathrm{RJ,QP}} = \frac{RJ}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right)\left(\frac{1}{n}\right)\max\left(1, \frac{QP}{BV^2}\right)}} \tag{8}$$

$$Z_{\mathrm{log,TP}} = \frac{\log(RV) - \log(BV)}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right)\left(\frac{1}{n}\right)\max\left(1, \frac{TP}{BV^2}\right)}} \tag{9}$$

$$Z_{\mathrm{log,QP}} = \frac{\log(RV) - \log(BV)}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right)\left(\frac{1}{n}\right)\max\left(1, \frac{QP}{BV^2}\right)}} \tag{10}$$

where the *relative jump* statistic RJ is defined as $\frac{RV-BV}{RV}$. Among these, Huang and Tauchen (2005) identify $Z_{\mathrm{RJ,TP}}$ as the statistic with the best finite-sample properties. We test the null hypothesis that a sample period contains no jumps using a standard z-test; if the test statistic Z exceeds the critical value $\Phi^{-1}(\alpha)$, where Φ is the standard normal distribution function, then we reject the hypothesis of no jumps at the α confidence level.

3.2 Jiang-Oomen Tests

Jiang and Oomen (2008) propose an alternative jump detection scheme. The test relies on a statistic they called *Swap Variance*, given by:

$$SwV = 2\sum_{i=2}^{n}(R_i - r_i) \tag{11}$$

where r_i is the geometric return defined before and R_i is the arithmetic return $\frac{P(t_i)-P(t_{i-1})}{P(t_{i-1})}$. $P(t_i)$ is defined as the price at time t_i, and thus $P(t_i) = \exp(p(t_i))$.

Jiang and Oomen then use the difference between Swap Variance and Realized Variance as the basis of their test statistic. Effectively, this exploits the higher-order moments of returns to identify discontinuous movements in the price series:

$$SwV - RV = \frac{1}{3}\sum_{i=2}^{n} r_i^3 + \frac{1}{12}\sum_{i=2}^{n} r_i^4 + \ldots + \frac{1}{2 \cdot k!}\sum_{i=2}^{n} r_i^{k+1} + \ldots \tag{12}$$

With a fine enough sampling frequency, jumps will cause a detectable increase in the value of this statistic, because high r_i values caused by discontinuities in the price process will be amplified by the larger exponents in the expansion.

To achieve a test statistic with an asymptotically standard normal distribution, we need to compute a scaling factor that depends on the *Integrated Sexticity*:

$$\Omega_{SwV} = n^2 \frac{\mu_6}{9} \frac{(n-1)\mu_{3/2}^{-4}}{n-4} \sum_{i=5}^{n} |r_i|^{3/2} |r_{i-1}|^{3/2} |r_{i-2}|^{3/2} |r_{i-3}|^{3/2} \tag{13}$$

From here, we can formulate several z-statistics that test the null hypothesis of no jumps in a sample period:

$$JO_{\text{Diff}} = \frac{n}{\sqrt{\Omega_{SwV}}}(SwV - RV) \tag{14}$$

$$JO_{\text{Log}} = \frac{nBV}{\sqrt{\Omega_{SwV}}}(\log(SwV) - \log(RV)) \tag{15}$$

$$JO_{\text{Ratio}} = \frac{nBV}{\sqrt{\Omega_{SwV}}}\left(1 - \frac{RV}{SwV}\right) \tag{16}$$

Jiang and Oomen provide Monte Carlo finite sample experiments that suggest that the JO_{Ratio} statistic is best. Again, we test the null hypothesis of no jumps using a standard z test.

3.3 Accounting for Market Microstructure Noise

Our data on stock prices is imperfect. It does not exactly reflect the fundamental values given by the models in Section 2; the market cannot possibly keep the current price in line with the theoretical price at all times. Instead, prices contain an element of *microstructure noise*. Mathematically, the observed log-price $p(t)$ is given by:

$$p(t) = p^*(t) + \epsilon_t \tag{17}$$

where ϵ_t represents a short-term deviation from the fundamental log-price $p^*(t)$.

This noise can distort the results of jump tests. In particular, if ϵ_t is independently and identically distributed (i.i.d), we will see negative serial correlation of returns. If an unusually high ϵ_t causes the observed price to display positive returns in one period, the price is more likely to decrease in the next period, as ϵ_{t+1} will probably be lower than ϵ_t. At sufficiently high frequencies, this negative correlation approaches -1 and dominates log-returns.

Anderson, Bollerslev, and Diebold (2004) suggest breaking this correlation by staggering returns: using only price data from every N minutes, even when observed prices are available at intervals of 1 minute or less. Huang and Tauchen (2005) confirm that this procedure makes jump tests more robust to microstructure noise. As a general approach to market microstructure noise, Anderson, Bollerslev, Diebold, and Labys (2000) recommend volatility signature plots, which display how the average realized variance corresponds to the sampling frequency. At small intervals, realized variance will be high due to microstructure noise, inflated by changes caused by the ϵ_t term. (In fact, if noise is i.i.d, realized variance will go to infinity as sampling becomes arbitrarily fine.) This effect will diminish as the sampling interval increases, and we can balance the objectives of robustness to microstructure noise and preserving the asymptotic properties of our estimators by choosing the interval where variance appears to stabilize. Using this technique, we will find that an interval of 15 minutes appears to be optimal for the stocks in our sample, and use it in our subsequent analysis. Figure 1 displays a volatility signature plot for one of the stocks in the sample.

4 The Effects of Dynamic Volatility: Theoretical Results

We first see whether there is any clear mathematical justification for why changes in volatility may cause bias in jump detection. Consider the expression for bipower variation over finitely spaced price data:

$$BV = \mu_1^{-2} \left(\frac{n-1}{n-2} \right) \sum_{i=3}^{n} |r_i||r_{i-1}| \tag{18}$$

Say that the sequence of geometric returns is generated by a diffusive process with deterministic volatility $\sigma(t)$ and zero drift. Then we can decompose r_i into the product $\sigma_i Z_i$, where Z_i is a standard normal random variable and $\sigma_i = \int_{t_{i-1}}^{t_i} \sigma(s)ds$. We now find:

$$BV = \mu_1^{-2} \left(\frac{n-1}{n-2} \right) \sum_{i=3}^{n} |Z_i||Z_{i-1}|\sigma_i \sigma_{i-1} \tag{19}$$

Recalling that $E[|Z_i|] = \mu_1$ by definition, taking expectations we find:

$$E[BV] = \mu_1^{-2} \left(\frac{n-1}{n-2} \right) \sum_{i=3}^{n} E[|Z_i||Z_{i-1}|]\sigma_i \sigma_{i-1} \tag{20}$$

$$= \frac{n-1}{n-2} \sum_{i=3}^{n} \sigma_i \sigma_{i-1} \tag{21}$$

Now, noting that $\sigma_i \sigma_{i-1} \leq \frac{\sigma_i^2 + \sigma_{i-1}^2}{2}$, we find:

$$E[BV] = \frac{n-1}{n-2} \sum_{i=3}^{n} \sigma_i \sigma_{i-1} \tag{22}$$

$$\leq \frac{n-1}{n-2} \left(\frac{\sigma_2^2 + \sigma_n^2}{2} + \sum_{i=3}^{n-1} \sigma_i^2 \right) \tag{23}$$

$$= \frac{n-1}{n-2} \left(E[RV] - \frac{\sigma_2^2 + \sigma_n^2}{2} \right) \tag{24}$$

$$= E[RV] + \frac{n-1}{n-2} \left(\frac{E[RV]}{n-1} - \frac{\sigma_2^2 + \sigma_n^2}{2} \right) \tag{25}$$

Equality only holds when all σ_i are the same, and the inequality is more pronounced as variation in σ_i increases. In particular, the difference between $\sigma_i \sigma_{i-1}$ and $\frac{\sigma_i^2 + \sigma_{i-1}^2}{2}$ is directly related to the magnitude of the difference between σ_i and σ_{i-1}:

$$\frac{\sigma_i^2 + \sigma_{i-1}^2}{2} - \sigma_i \sigma_{i-1} = \frac{(\sigma_i - \sigma_{i-1})^2}{2} \tag{26}$$

From equation 25, we observe that if the σ_i are not all the same, the expected value of BV is less than the expected value of RV plus a boundary term. The boundary term will be positive if the average of square volatility in the first and last periods is less than the average of square volatility by period throughout the sample, and negative otherwise. Thus, if volatility tends to be higher at the beginning and end of a day, the gap between expected BV and RV over that day will be even larger than the first inequality suggests.

Recall that RV and BV are both estimators of integrated variance, and in a diffusive price process, they will asymptotically give the same quantity. Still, as long as volatility is not constant and the boundary term is not too positive, in finite samples the expected value of BV will be less than the expected value of RV, skewing the distribution of the BNS test statistics and biasing the test in favor of finding jumps where none actually exist.

Using the same inequality argument, we can conclude that our jump-robust estimators of quarticity and sexticity, which rely on multiplying consecutive geometric returns to dampen the effect of a jump in one period, will also be biased downward and affected by a similar boundary term. This may further damage the BNS test statistic, since either tripower quarticity or quadpower quarticity is used in the denominator. The effect on the denominator, however, is not clear, because in the forms we examine it also contains a BV^2 term, which will also be downwardly biased and may cancel out some or all of the downward bias from TP or QP.

Similarly, we may expect problems with the Jiang-Oomen statistics, although the exact effect is again unclear. In all cases, the Ω_{SwV} statistic in the denominator, which relies on a sexticity estimate, is likely to be biased downward. With the log and ratio statistics, however, downward bias of the BV term in the numerator may counteract this effect, leaving the overall change ambiguous. To determine the comparative magnitudes of these biases and their cumulative effect, it is necessary to identify the sources of dynamic volatility and compute the effects on simulated data.

5 Intraday Pattern in Volatility

Although volatility is generally considered to be a long-memory process, *spot volatility* undergoes a dramatic pattern during most days: it starts high, moves down by a factor of two or more to a midday trough, and then moderately increases before the end of the day. This pattern is remarkably widespread and has been repeatedly confirmed throughout the literature—for instance, by Wood, McInish, and Ord (1985), Lockwood and Linn (1990), and Andersen and Bollerslev (1997). Figure 2 displays the average absolute geometric return for each minute during the trading day for a sample of four securities. Observe that the patterns are remarkably similar. For instance, they all include a noticeable jump in volatility around 25 minutes into the trading day (roughly 10:00), which presumably occurs because 10 AM[3] is a common time for announcements.

We found in the previous section that rapid changes in volatility during the sample period will bias several statistics (BV, TP, QP, and Ω_{SwV}) downward, particularly if boundary terms like the one in Equation 25 are negative. Figure 2 makes clear that these terms will indeed be negative: volatility is higher at the beginning and end of each day than in the sample as a whole. Although this pattern is not the only dynamic aspect of spot volatility, it is both easily estimated and possibly dominant over short sample periods.

[3]As we will discuss briefly in the data section, we truncate the first five minutes of the trading day; hence the 25th minute of our daily data is 10:00

6 Measuring the Effect of Dynamic Volatility

To investigate the effect of these large intraday swings in volatility, we carry out simple simulations of a diffusive price process. Our goal is to examine the behavior of a jump tests on a sample where there are no jumps, to provide evidence of the number of spurious jumps they detect. Recall Equation 2 for a general price process with jumps:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t)$$

Since we are simulating a diffusive price process, we set $\kappa(t) = 0$. Jump statistics are typically applied to individual days, and the $\mu(t)$ drift term is therefore insignificant enough to ignore in our simulation as well. We are therefore left with simply:

$$dp(t) = \sigma(t)dW(t) \tag{27}$$

We break up the day from 9:35 AM to 4:00 PM into 384 one-minute intervals and simulate using the Euler scheme (for $i = 1 \ldots 385$):

$$p(t_{i+1}) - p(t_i) = \sigma_i Z \qquad Z \sim N(0,1) \tag{28}$$

In one simulation, volatility is assumed to be constant throughout the day: $\sigma_i = c$. In another, it is determined by the minute-by-minute volatility pattern of the stock in question: σ_i set equal to the average absolute return $|p(t_{i-1}) - p(t_i)|$ over all days in the sample.

As is evident from above, neither simulation incorporates any other dynamics—leverage effects, feedbacks, or jumps—into the volatility process. The geometric returns at each minute are randomly pulled from a normal distribution, with constant standard deviation in the first case and standard deviation scaled in the second case by the intraday volatility pattern. If the jump statistics are accurate, when applied to such a simple price process they should produce roughly the normal aymptotic distribution suggested by theory, which we use in practice to set critical values. Our goal is to see whether this actually holds in the presence of intraday volatility dynamics. If the distribution we obtain is not as expected, we investigate whether its failure is enough to account for a sizable fraction of the jumps detected in applications of the jump tests to the observed data. Specifically, we compare the number of jumps detected in jump-free simulated data with the pattern to the number of jumps detected in observed data. If the former is a sizable percentage of the latter, we may conclude that many of the jumps identified in observed data result from flaws in the tests rather than actual discontinuous movements in prices.

For each stock in the sample, we generate 100,000 "trading days," with 385 minutes each, of simulated data for both types of volatility process. Note that while the simulated process with constant volatility does not depend on the stock chosen, the simulated process with an intraday volatility pattern depends on the pattern in the data for the respective stock. We apply all seven tests listed earlier—four variants of the BNS test and three Jiang-Oomen tests—to examine each day for the presence of a jump. We also apply the tests directly to the observed data, and compare the fraction of jumps detected with the fraction in the simulated data. In the process, we maintain data on the components of the jump tests: the average values of RV, BV, and other estimators, along with the test statistics themselves. To see how our results change with different sampling frequencies, we repeat the jump tests at all sampling intervals from length 1 minute to 30 minutes. For each stock, all tests are performed on the same simulated (and observed) pricing data.

7 Data

The stock whose results we examine in most detail will be JP Morgan (JPM), which displays intraday volatility behavior representative of high-volume securities in general. We also perform our tests on simulations using data from a sample of stocks chosen for their breadth, high volume, and high market capitalization: Coca-Cola (KO), Exxon Mobil (XOM), Intel (INTC), Microsoft (MSFT), and Wal-Mart (WMT).

The price data were obtained from price-data.com, a commercial data vendor, and include every minute from 9:35 AM to 4:00 PM on trading days from 1997 to early 2009. The number of trading days actually included in the sample for each stock ranges from 2264 to 2924 days; the smallest value comes from Exxon Mobil, which is only recorded after its 1999 merger.

8 Results

Table 1 displays the results from running several jump tests on diffusively simulated data using 15-minute intervals, and how these jump frequencies compare to the fraction of jumps detected in the observed price data. We find that the relative jump BNS test using tripower quarticity is the least likely to produce Type I error, which is consistent with the Monte Carlo results of Huang and Tauchen. As predicted by our earlier theoretical discussion, the simulation that includes an intraday volatility pattern has significantly higher likelihood of Type I error. Under most tests, the fraction of spurious jumps detected in the simulated data is half or more of the fraction detected in the actual data, suggesting that many of the jumps identified in empirical studies that use these tests are spurious. In particular, with the $Z_{\mathrm{RJ,TP,0.99}}$ statistic, the number of "jumps" in the jump-free simulated data is approximately 55% of the number of jumps detected in observed data.

Table 12 illustrates how the different components of the BNS and Jiang Oomen jump tests are affected by the introduction of dynamic volatility. We compare the mean estimates from our simulated data to the true expected values of these components, which are easily estimated from the volatility process used in our simulation. As predicted, since our simulation consists of diffusive price movements, the realized variance statistic is nearly unbiased for integrated volatility in both cases, with a very small mean percentage error. Bipower variation, however, is strikingly affected by the introduction of intraday volatility patterns: while it is unbiased with static volatility, after the pattern is added it displays a bias of nearly negative 4 percent. For comparison, in the observed data, total BV is 6.45% less than total RV. Meanwhile, the quarticity estimates are even more significantly affected, as TP and QP are biased downward by 19% and 26%, respectively.

We can see how this biased estimation affects the sample distribution of z-scores in Figure 2, which shows superimposed kernel density plots of z-scores both with and without the intraday pattern included in the simulation. The intraday volatility dynamics cause the z-score distribution to widen, as negatively biased quarticity estimates make the denominator of the statistic smaller, and move to the right, as negatively biased bipower variation leads the numerator to have a mean greater than zero. Figure 4 illustrates the increasing negative bias of BV as the length of the sampling interval increases, making clear that the level of bias hinges on the choice of sampling frequency. This is consistent with our theoretical results.

Returning to table 1, we observe that an even higher fraction of jumps detected by the Jiang Oomen tests can be explained by a diffusive price process following an intraday volatility pattern. Consistent with the results of Jiang and Oomen (2008), we find that the JO_{ratio} test is least likely to produce Type 1 error. Still, the fraction of jump days detected by this test on the scaled, simulated data is almost two-thirds of that detected in the actual data, implying that only a minority of "jump" days detected in the observed data are likely to contain actual, statistically significant discontinuous movements in price.

In table 3, we observe results for several more stocks. The table displays the BNS and Jiang Oomen test variants least likely to produce Type I error, and identified in the literature for the best finite sample properties, $Z_{\text{RJ,TP}}$ and JO_{ratio}. The tests are performed on simulated data scaled by the volatility pattern of each respective stock, and on the observed data from each stock itself. Our results are broadly consistent for all stocks: again, for the BNS test, the fraction of jumps identified in the simulated data is roughly half the fraction in the actual data, and for the Jiang Oomen test the ratio is roughly two-thirds.

9 Possible Remedies

It may be possible to partially remedy the issues raised in this paper. As we found in the theoretical discussion and substantiated with Monte Carlo simulations, bias in jump tests arises directly from volatility levels that differ between consecutive sampling intervals. Hence, if we can implement tests such that consecutive intervals have similar volatility levels, we will dramatically limit this bias. Two strategies are apparent:

1. Before applying the jump tests, scale log-returns in each interval by the intraday pattern so that average volatility levels in each interval are approximately equal.

2. Drop the requirement that sampling intervals must be equispaced. Instead, set intervals so that the average volatility in each interval over the sample is approximately equal.

First we implement strategy 1. If $r_{i,j}$ is the log-return in period i on day j (where the period i may correspond to a sampling interval of arbitrary length), and there are m days total, we replace it with the modified return:

$$\tilde{r}_{i,j} = \frac{r_{i,j}}{\frac{1}{m-1}\sum_{j' \neq j} |r_{i,j}|} \tag{29}$$

Note that since $r_{i,j}$ may include a jump component, we do not include it in our calculation of the average log-return in the denominator.

We then run the BNS jump tests on this modified data. (We do not attempt this strategy with the Jiang-Oomen tests, since it alters the scale of the log-returns. This is irrelevant for the BNS tests, which are scale-invariant with respect to log-returns, but it is problematic for the Jiang-Oomen tests, which are not.) The results for the $Z_{\mathrm{RJ,TP,0.99}}$ test, identified in both this paper and in Huang and Tauchen (2005) as the best-behaved jump test at the 0.99 significance level, are displayed in the second row of Table 4 for JPM. We keep the same 15-minute intervals used in our earlier tests.

The fraction of days flagged as jumps by this test declines from .0585 using the unmodified data to .0352. The table also shows the average value of $\frac{RV-BV}{RV}$ over all days in the sample, where RV and BV are calculated from the modified log-returns. This indicates the jump component's fraction of total "variation" in the modified returns. This is .0292, versus .0645 for the unmodified returns. Clearly, scaling by the intraday volatility pattern dramatically lowers the apparent importance of jumps.

One possible weakness with this strategy is that if jumps do not follow the same intraday pattern as the diffusive component of price movement, the denominator in (29) will be greater during times in the day when jumps are more common. To address these concerns, we may scale by the square root of average local bipower variation, which is relatively more robust to jumps. In this case, we take:

$$\tilde{r}_{i,j} = \begin{cases} \frac{r_{1,j}}{\sqrt{\frac{1}{m}\sum_j |r_{1,j}||r_{2,j}|}} & i = 1 \\ \frac{r_{i,j}}{\sqrt{\frac{1}{2m}\left(\sum_j |r_{i-1,j}||r_{i,j}|+\sum_j |r_{i,j}||r_{i+1,j}|\right)}} & 1 < i < c \\ \frac{r_{c,j}}{\sqrt{\frac{1}{m}\sum_j |r_{c-1,j}||r_{c,j}|}} & i = c \end{cases} \tag{30}$$

where c is the index of the final return interval of the day.

The results of this modification are displayed in the third row of Table 4. They are similar to those from before: the $Z_{\mathrm{RJ,TP,0.99}}$ test indicates that .0359 of days contain jumps, and gives an average value of .0324 for $\frac{RV-BV}{RV}$ over the full sample. These values are still substantially below those obtained using the unmodified data.

We also implement strategy 2 from above. Letting $r_{i,j} = p_{i+1,j} - p_{i,j}$, where $p_{i,j}$ is the log-price in the ith minute on the jth day, we define:

$$\overline{r_i^2} = \frac{1}{m}\sum_j r_{i,j}^2 \tag{31}$$

$$\overline{RV} = \sum_i \overline{r_i^2} \tag{32}$$

Now we define

$$g(i) = \sum_{i'=1}^{i} \overline{r_{i'}^2} \tag{33}$$

If we aim to split the day into n sampling intervals, then for $k = 0, \ldots, n$ we write:

$$d_k = \min\{i;\ g(i) \geq \frac{k}{n}\overline{RV}\} \tag{34}$$

Then we define the kth sampling interval as the interval from d_{k-1} to d_k and the kth sample log-return on day j as

$$\tilde{r}_{k,j} = p_{d_k,j} - p_{d_{k-1},j} \tag{35}$$

By construction, the average squared return in each of these intervals will be approximately $\frac{k}{n}\overline{RV}$.

Applying the $Z_{\text{RJ,TP,0.99}}$ and $JO_{\text{ratio,0.99}}$ tests to the log-returns $\tilde{r}_{k,j}$ produced following this strategy, and splitting the day into the same number of sampling intervals we had with 15-minute sampling, we obtain the results in the fourth row of Table 4. The fraction of days flagged by $Z_{\text{RJ,TP,0.99}}$ is now .0387, and the fraction of days flagged by $JO_{\text{ratio,0.99}}$ is .0698, down from .0585 and .1040, respectively, using the unmodified tests. The full-sample daily average of $\frac{RV-BV}{RV}$ is .0311.

While we are specifying the intervals so that average RV is roughly constant across all intervals, RV includes both the diffusive and jump components of log-price variation, and the theory in Section 4 only suggests that we should try to make the *diffusive* component constant across all intervals. Hence we will also specify intervals by attempting to equate total BV in each interval. Letting d be the index of the final minute in the day, we write:

$$\overline{BV}_i = \begin{cases} \frac{1}{m}\sum_j |r_{1,j}||r_{2,j}| & i = 1 \\ \frac{1}{2m}\left(\sum_j |r_{i-1,j}||r_{i,j}| + \sum_j |r_{i,j}||r_{i+1,j}|\right) & 1 < i < d \\ \frac{1}{m}\sum_j |r_{d-1,j}||r_{d,j}| & i = d \end{cases} \tag{36}$$

$$\overline{BV} = \sum_i \overline{BV}_i \tag{37}$$

Now, like before, we define:

$$g(i) = \sum_{i'=1}^{i} \overline{BV}_{i'} \tag{38}$$

For $k = 0, \ldots, n$ we write:

$$d_k = \min\{i;\ g(i) \geq \frac{k}{n}\overline{BV}\} \tag{39}$$

Again, we define the kth sampling interval as the interval from d_{k-1} to d_k and the kth sample log-return on day j as

$$\tilde{r}_{k,j} = p_{d_k,j} - p_{d_{k-1},j} \tag{40}$$

By construction, the average bipower variation at each of these intervals will be approximately $\frac{k}{n}\overline{BV}$.

Applying tests to the new log-returns $\tilde{r}_{k,j}$, we obtain the results in the fifth row of Table 4. The $Z_{\mathrm{RJ,TP,0.99}}$ test identifies .0325 of days as containing jumps, and $JO_{\mathrm{ratio,0.99}}$ flags .0694. The full-sample average of $\frac{Rv-BV}{RV}$ declines to .0276.

Although the four jump test modifications outlined in this section produce slightly different results, they are remarkably similar in comparison to the unmodified jump tests. The new estimates for the fraction of jump days identified by the $Z_{\mathrm{RJ,TP,0.99}}$ test range from .0325 to .0387, substantially lower than the .0585 obtained using the original tests. The full-sample average of $\frac{RV-BV}{RV}$ ranges from .0276 to .0324, compared to our earlier value of .0645. Finally, although we were only able to apply the last two modifications to the $JO_{\mathrm{ratio,0.99}}$ test, they also display a marked decline, from .1040 to .0698 and .0694.

These results substantiate our earlier finding that the intraday volatility pattern inflates the apparent importance of jumps in the sample: when its effect is minimized through various modifications to the tests, the prevalence of jumps declines in a dramatic and consistent way.

10 Conclusion

Motivated by recent research that identifies incoherent results from common jump tests, we demonstrate that the finite sample properties of these tests are not robust to large intraday swings in volatility. At a sampling interval (15 minutes) chosen to limit the effects of market microstructure noise in practical applications, Monte Carlo simulations reveal that a pure diffusion process scaled by the intraday volatility pattern produces data with one-half to two-thirds the fraction of jump days—depending on the test—obtained from the actual data. The full-sample difference between realized variance and bipower variation in the simulation, often interpreted as an indicator of the size of the jump component of volatility, is also more than half its value in the real data. Existing estimates of jumps that rely on these tests are therefore likely to be significantly overstated.

It is important to note that the weaknesses of our jump tests are not the consequence of intraday volatility patterns per se, but rather of a volatility process that displays intraday swings of sufficient magnitude. Any additional changes in volatility will cause further distortion to the size of our test statistics; as a result, our results likely represent a lower bound on the extent to which currently detected jumps are spurious. This poses a serious problem for any attempt to determine whether outliers in price movement data arise from complicated volatility dynamics or jumps in the price process itself. As we have seen, a higher rate of change in the volatility process directly inflates the fraction of spurious jumps detected in the data, and our current set of tools is therefore liable to confuse the two phenomena.

Nevertheless, we continue by providing several intuitive, practical modifications to jump tests that seek to minimize the effects of the intraday volatility pattern. Two modifications change the scale of the log-returns themselves before applying the tests, while two other modifications move from equispaced intervals to a sampling scheme that roughly equalizes the volatility level across all intervals. The results are in line with the implications of our earlier work: once the modifications are made, the measured frequency and importance of jumps drop dramatically. The similarity of the results from different methods to modify jump tests is particularly striking, and it suggests that intraday volatility changes are indeed inflating the results of conventional jump detection.

Since they only correct for the intraday pattern, and not all sources of dynamic volatility, our adjustments cannot fully eliminate the problem of spurious jump detection. They are also dependent on knowledge of the form of the intraday volatility pattern, which may change over time. One possible future approach is to sample according to some proxy for volatility. For instance, we may abandon fixed intervals in "calendar time" altogether and instead use the cumulative number of transactions as our measure of time. To the extent that the frequency of transactions reflects spot volatility, this approach will produce sampling intervals of roughly similar integrated variance. Although this technique is rare in the existing literature on jump tests, Oomen (2005) applies it to estimation of realized variance and finds that it results in substantially improved accuracy. Its possible application to jump detection merits careful inquiry.

Overall, our results emphasize the importance of looking beyond the asymptotic properties of statistical tests to determine the reality of finite sample application. While the components of jump tests are consistent estimators of integrated variance and quarticity, they suffer from a simple bias in finite samples that has severe consequences for the reliability of jump detection. Despite the impressive high-frequency data at our disposal, financial markets are not yet liquid enough for asymptotic results to carry much practical meaning, as microstructure noise forces us to limit our sampling frequency to less impressive levels.

11 Figures

Figure 1: Signature Volatility Plot for JPM

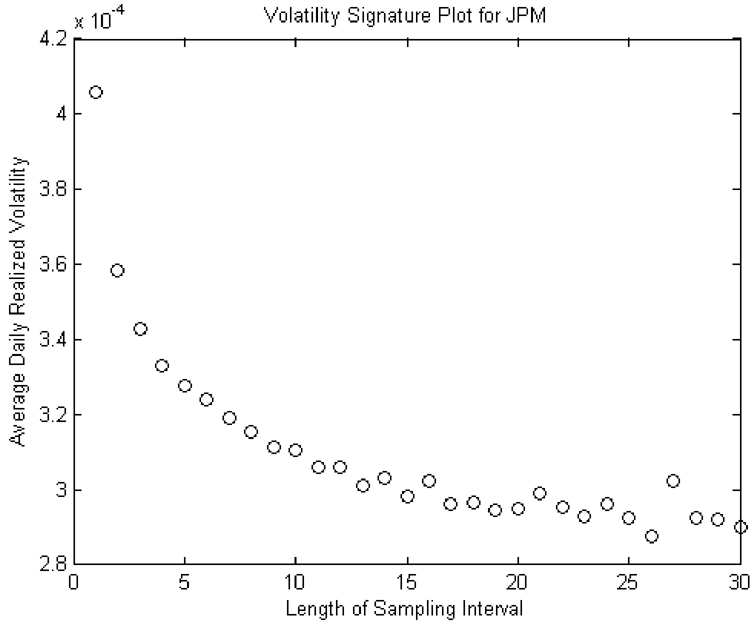

Figure 1 is a Signature Volatility Plot for the JPM data. It shows the average daily realized variance when realized variance is calculated at every sampling interval from 1 to 30. Realized variance is noticeably higher when the sampling interval is small; as discussed in the paper, this is the result of market microstructure noise, which causes realized variation to spike as the sampling interval approaches zero.

Effectively, the level of realized variance is used as a proxy for the degree to which microstructure noise affects statistics calculated at sampling intervals of various length. Through visual inspection, we select a sampling interval at which the realized variance has stopped decreasing significantly. For JPM and the other stocks in our sample, this interval was chosen to be 15 minutes. Though most calculations in the paper are done at a variety of sampling intervals, 15 minutes is used as the default length when a single result is presented.

Figure 2: Intraday Volatility Pattern for a Sample of Stocks

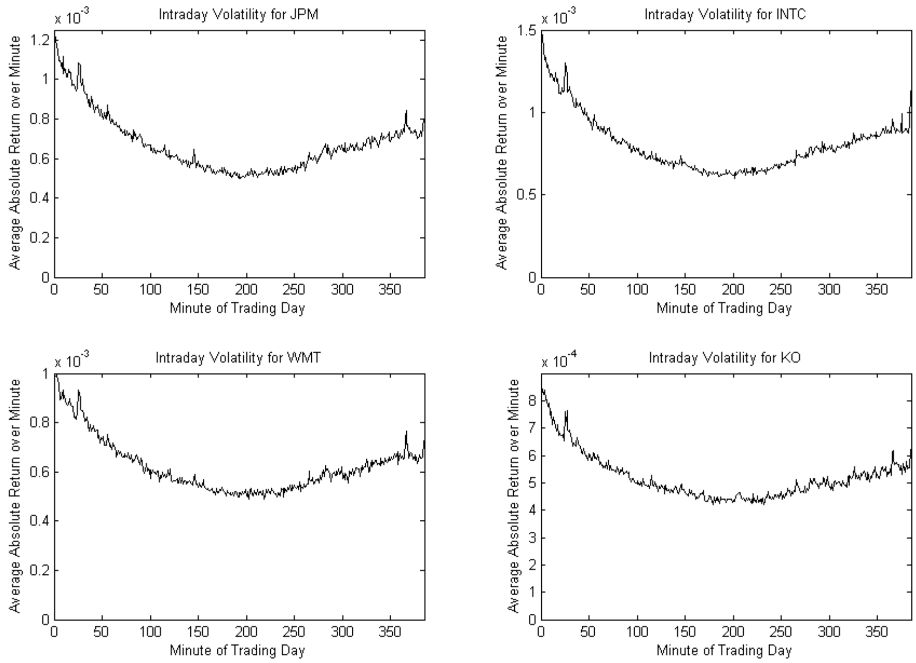

Figure 2 shows the minute-by-minute pattern in average absolute returns for four representative stocks: JP Morgan, Intel, Wal-Mart, and Coca-Cola. This pattern is remarkably similar in all four cases, and is consistent with results in the literature about intraday volatility patterns. Average volatility starts highest at the beginning of the day, falls to a midday trough of roughly one-half the maximum volatility, and gradually recovers during the rest of the day.

Figure 3: Kernel Density Estimates for $Z_{\mathrm{RJ,TP}}$ Test Statistic With and Without Intraday Pattern (15-minute sampling)

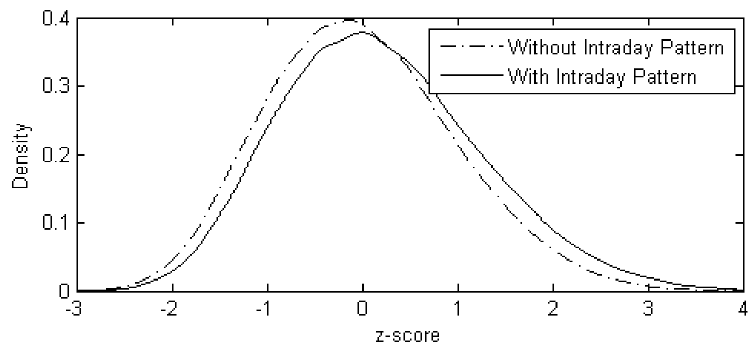

Figure 3 shows kernel density estimates for the distribution of the $Z_{\mathrm{RJ,TP}}$ test statistic when it is applied to simulated data generated with volatility scaled by an intraday volatility pattern and when it is applied to simulated data generated with constant volatility throughout the day. Some clear differences are evident between the two densities: the density in the "with pattern" case is wider and shifted to the right. This agrees with our theoretical discussion: the negative bias in BV moves the distribution of Z scores to the right, while the negative bias in TP widens the distribution. Observe that the "without pattern" density much more closely resembles a Gaussian, suggesting that the asymptotic normality of the $Z_{\mathrm{RJ,TP}}$ is more nearly achieved when volatility is constant throughout the day.

For this figure, the intraday volatility pattern used for the simulation is taken from the JPM data, and the $Z_{\mathrm{RJ,TP}}$ test statistic is applied with a sampling frequency of 15 minutes.

Figure 4: Negative Bias of Bipower Variation as an Estimator of Integrated Variance: Simulated Data with Intraday Pattern

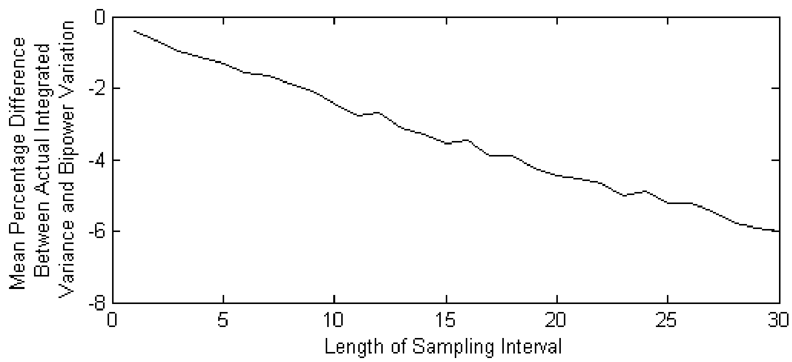

Figure 4 shows the negative bias of bipower variation as an estimator for integrated variance as a function of the sampling interval used to calculate bipower variation. In this case, we examine the bias for data simulated with the intraday volatility pattern for JPM. Since we are simulating the data, we know the "true" integrated variance, and the mean percentage difference is calculated as $\frac{BV-IV}{IV}$, where BV is the mean bipower variation and IV is the true integrated variance.

Observe that the negative bias of bipower variation steadily worsens as the sampling interval becomes larger. This substantiates our theoretical discussion, which showed that the bias in bipower variation is roughly proportional to the sum of squared differences between volatility in consecutive sampling intervals. When there is a consistent intraday volatility pattern, the difference in volatility levels between two consecutive sampling intervals is roughly proportional to the length of the sampling intervals, implying that the squared differences are proportional to the squared length. The number of sampling intervals decreases in proportion to the length, and altogether the sum of squared differences—which we showed produces the bias—should be roughly proportional to the sample length, exactly as this figure depicts.

12 Tables

All tables display statistics that use 15-minute sampling intervals.

Table 1: Fraction of Days Flagged as Jumps: JPM

Test	Simulated, No Pattern	Simulated With Pattern	Actual Data
$Z_{\mathrm{RJ,TP},0.99}$	.0144	.0328	.0585
$Z_{\mathrm{RJ,TP},0.999}$	.0018	.0073	.0151
$Z_{\mathrm{RJ,QP},0.99}$	.0149	.0347	.0626
$Z_{\mathrm{RJ,QP},0.999}$	.0021	.0073	.0188
$Z_{\mathrm{log,TP},0.99}$	.0329	.0621	.0975
$Z_{\mathrm{log,TP},0.999}$	.0110	.0270	.0520
$Z_{\mathrm{log,QP},0.99}$	.0334	.0639	.1023
$Z_{\mathrm{log,QP},0.999}$	.0114	.0284	.0551
$JO_{\mathrm{diff},0.99}$	.0616	.0965	.1314
$JO_{\mathrm{diff},0.999}$	.0319	.0604	.0958
$JO_{\mathrm{log},0.99}$	.0457	.0761	.1102
$JO_{\mathrm{log},0.999}$	.0181	.0400	.0626
$JO_{\mathrm{ratio},0.99}$	.0446	.0761	.1040
$JO_{\mathrm{ratio},0.999}$	.0174	.0400	.0626

Table 1 displays results for 14 jump tests, which are derived from 7 jump tests and 2 significance levels for each test. Each row corresponds to a jump test. The first column shows the fraction of days flagged as jumps when the tests are applied to simulated returns with constant volatility throughout the day. The second column shows the fraction of days flagged as jumps when tests are applied to simulated returns, where these returns are simulated following an intraday volatility pattern taken from the JPM data. The third column shows the fraction of days flagged as jumps when tests are applied to the observed price data for JPM.

In every row, the third column has the highest fraction and the first column has the lowest: jumps are least frequently detected in simulated returns with constant volatility, and most frequently detected in observed re turns. Observe that for the Z-statistics, the second column is in most cases at least half the third column. In other words, jumps are detected in data simulated without jumps (but with an intraday volatility pattern) at at least half the rate they are detected in the observed data. For the JO-statistics, this ratio increases to as much as two-thirds.

Table 2: Percentage Difference Between Average Daily Sample Statistics, from Simulated Data, and True Values of Estimated Quantities: JPM

Statistic	Without Intraday Pattern	With Intraday Pattern
RV	0.11%	0.09%
BV	0.30%	-3.77%
TP	0.45%	-18.82%
QP	0.62%	-25.51%

Table 2 displays the measured bias in four sample statistics that are important components of jump tests: realized variance, bipower variation, tripower variation, and quadpower variation. This bias is shown for two kinds of simulated data: data simulated with constant volatility throughout the day ("without intraday pattern") and data simulated using volatilities from the intraday pattern displayed by JPM ("with intraday pattern"). Since we are simulating the data, we know the true underlying integrated variance and integrated quarticity, and we record here the mean percentage error with which RV, BV, TP, and QP measure these values. For the mean error for RV is calculated as $\frac{RV-IV}{IV}$, where RV is the mean calculated realized variance and IV is the true integrated variance.

All statistics are unbiased when applied to data that is simulated without an intraday pattern. When data is simulated with the pattern, however, only RV is unbiased. BV is biased downward by slightly less than 4 percent, and TP and QP are biased downward by roughly one fifth and one quarter, respectively. These statistics are all calculated using 15 minute sampling intervals.

Table 3: Fraction of Jump Days Detected in Simulated Data with Intraday Pattern and in Observed Data, for Various Stocks

Stock	$Z_{\mathrm{RJ,TP},0.99}$	$JO_{\mathrm{ratio},0.99}$
INTC (simulated)	.0302	.0669
INTC (observed)	.0510	.0970
JPM (simulated)	.0328	.0761
JPM (observed)	.0585	.1040
KO (simulated)	.0267	.0568
KO (observed)	.0578	.0971
MSFT (simulated)	.0301	.0634
MSFT (observed)	.0572	.0979
WMT (simulated)	.0262	.0600
WMT (observed)	.0592	.0904
XOM (simulated)	.0258	.0578
XOM (observed)	.0477	.0888

Table 3 displays results from applying two jump tests to simulated and actual data for several different stocks. The "simulated" data for each stock is the set of returns simulated using the intraday volatility pattern drawn from that stock's observed returns. The "observed" data for each stock is simply that stock's set of observed returns in the sample. The two jump tests shown are the variants of the BNS test and the Jiang-Oomen test that Huang and Tauchen (2005) and Jiang and Oomen (2008), respectively, identify as having the best finite sample properties. All tests are performed at the 0.99 significance level and with 15-minute sampling.

Table 4: Jump Test Results Under Various Modifications: JPM

Modification	$Z_{\mathrm{RJ,TP},0.99}$	$JO_{\mathrm{ratio},0.99}$	$(RV - BV)/RV$ (full sample)
None	.0585	.1040	.0645
Scaled Returns (1)	.0352	X	.0292
Scaled Returns (2)	.0359	X	.0324
Non-Equispaced Intervals (1)	.0387	.0698	.0311
Non-Equispaced Intervals (2)	.0325	.0694	.0276

Table 4 displays the fraction of days flagged as containing jumps by the $Z_{\mathrm{RJ,TP},0.99}$ and $JO_{\mathrm{ratio},0.99}$ tests, in addition to the average of $\frac{RV-BV}{RV}$ over all days in the sample, under several proposed modifications to the log-returns used in the tests. The original results are included in the "None" row, and the results from the first and second methods of scaling returns in Section 9 are included in the "Scaled Returns (1)" and "Scaled Returns (2)" rows, respectively. The Xs indicate that a test cannot be applied under the modification. Finally, the results from the first and second methods of altering sampling intervals in Section 9 are included in the "Non-Equispaced Intervals (1)" and "Non-Equispaced Intervals (2)" rows, respectively.

References

[1] Y. Ait-Sahalia and J. Jacod. Testing for Jumps in a Discretely Observed Process. *Annals of Statistics*, 35(1):355–392, 2007.

[2] T. Andersen and L. Benzoni. Realized Volatility. Federal Reserve Bank of Chicago: Working Paper WP-08-14, 2008.

[3] T. Andersen, L. Benzoni, and J. Lund. An Empirical Investigation of Continuous-Time Equity Return Models. *Journal of Finance*, 57(3):1239–1284, 2002.

[4] T. Andersen and T. Bollerslev. Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance*, 4(2–3):115–158, 1997.

[5] T. Andersen, T. Bollerslev, and F. Diebold. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Review of Economics and Statistics*, 89(4):701–720, 2007.

[6] O. Barndorff-Nielsen and N. Shephard. Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, 2(1):1–37, 2004.

[7] O. Barndorff-Nielsen and N. Shephard. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. *Journal of Financial Econometrics*, 4(1):1–30, 2006.

[8] F. Drost, T. Nijman, and B. Werker. Estimation and Testing in Models Containing Both Jumps and Conditional Heteroskedasticity. *Journal of Business and Economic Statistics*, 16(2):237–243, 1998.

[9] X. Huang and G. Tauchen. The Relative Contribution of Jumps to Total Price Variance. *Journal of Financial Econometrics*, 3(4):456–499, 2005.

[10] G. Jiang and R. Oomen. Testing for Jumps When Asset Prices Are Observed with Noise: a Swap Variance Approach. *Journal of Econometrics*, 144(2):352–370, 2008.

[11] S. Kou. A Jump-Diffusion Model for Option Pricing. *Management Science*, 48(8):1086–1101, 2002.

[12] S. Lee and P. Mykland. Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies*, 21(6):2535–2563, 2008.

[13] L. Lockwood and S. Linn. An Examination of Stock Market Return Volatility During Overnight and Intraday Periods. *Journal of Finance*, 45(2):591–601, 1990.

[14] J. Maheu and T. McCurdy. News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns. *Journal of Finance*, 59(2):755–793, 2004.

[15] R. Merton. Option Pricing When Underlying Stock Returns are Discontinuous. *Journal of Financial Economics*, 3(1–2):125–144, 1976.

[16] R. Oomen. Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes. *Journal of Financial Econometrics*, 3(4):555–577, 2005.

[17] M. Schwert. Hop, Skip, and Jump – What Are Modern "Jump" Tests Finding in Stock Returns? Duke University Senior Honors Thesis, 2009.

[18] P. Van Tassel. Patterns Within the Trading Day: Volatility and Jump Discontinuities in High-Frequency Equity Price Series. Duke University Senior Honors Thesis, 2009.

[19] R. Wood, T. McInish, and J. Ord. An Investigation of Transaction Data for NYSE Stocks. *Journal of Finance*, 40(3):723–739, 1985.