Cross-Stock Comparisons of the Relative Contribution of Jumps

Vivek Bhattacharya

Spring 2011

Economics 201FS

Honors Junior Workshop in Financial Econometrics

The Duke Community Standard was upheld in the completion of this report.
1 Introduction

The availability of high-frequency financial data over the past decade has spawned much work to carefully understand financial time series. An especially interesting empirical observation is that stock prices seem to include “jumps,” discontinuous movements in the price data often as a result of firm-specific news or macroeconomic events. Jumps can complicate standard options pricing and portfolio management problems, as many common models—such as the Black-Scholes model—are based on the assumption of a continuous, or “diffusive,” process. Given the presence of these jumps, it is informative to separate out the volatility in a particular trading day or week as that attributable to the diffusive process and that attributable to the jumps. It is interesting to determine how much of the total variance of a security is due to the jumps in the process. Statistics that utilize high-frequency data to estimate such quantities have been developed and studied extensively in the literature. We will provide an overview of these estimators before discussing our methodology and results.

This paper will focus on the time series of the relative contribution of jumps to the total variance. This time series has been studied in many contexts, such as by Barndorff-Nielsen and Shephard (2004) as the test statistic for a hypothesis test for the presence of intraday jumps and by Bollerslev, Kretschmer, Pigorsch, and Tauchen (2009) as an extension to heterogeneous autoregressive models for volatility. Despite the relatively high level of interest in this quantity in the literature, however, not much work has been done on studying the relationship between these time series across pairs of stocks. In the work conducted this semester, we take a first step towards addressing the relation between the relative contribution of jumps of different stocks. In particular, we spend most of the paper discussion the correlation between these time series. We also attempt to identify underlying properties of the price processes—such as diffusive volatility, the frequency of jumps, and other factors—that may affect this correlation.

Section 2 gives a general overview of the literature and presents relevant estimators as well as discussions of important empirical considerations, such as microstructure noise. Section 3 then presents the initial analysis that motivated the work this semester. We begin with a brief analysis of relative contribution of jump series for single stocks and then consider pairs of these time series simultaneously. In this section, we note some empirical regularities that we will explore. Section 4 offers a detailed description of the Monte Carlo setup we employ in this paper, and Section 5 summarizes the main results from the Monte Carlo studies. We offer some concluding remarks in Section 6 and suggest future directions.
2 Theoretical Background

2.1 Continuous Time Models for Returns

Throughout this paper, we will keep in mind a stochastic model of returns where an efficient price process \( p(t) \) given by

\[
dp(t) = \mu(t)dt + \sigma(t)dw(t),
\]

where \( \mu(t) \) and \( \sigma(t) \) give the drift and the volatility of the security, respectively, and \( dw(t) \) is a Wiener increment for the standard Brownian motion process \( w(t) \). Furthermore, we assume that \( \sigma(t) \) is independent of \( W(t) \). Note that \( p(t) \) is the log of the dollar price of the stock, and as a result Equation 1 is an assumption that the efficient price of a security follows a geometric Brownian motion process. To incorporate jumps into the return process, we simply introduce a new term \( dJ(t) \) to arrive at

\[
dp(t) = \mu(t)dt + \sigma(t)dw(t) + dJ(t),
\]

where \( J(t) = J(s) = \sum_{s \leq t' \leq t} \kappa(t') \) and \( \kappa(\cdot) \) is the magnitude of a jump. We often write \( dJ(t) = \kappa(t)dq(t) \) where \( q(t) \) is a (Poisson) counting process.

A goal of financial econometrics, as mentioned in the introduction, is to disentangle the jump component from the continuous component. A number of factors complicate this goal. First, we only observe the price at discrete time intervals—when the security is traded. Second, instead of observing the efficient price \( p(t) \), the econometrician observes the process \( p^*(t) \), which is contaminated with microstructure noise. We write

\[
p^*(t) = p(t) + \epsilon_t,
\]

where \( \epsilon_t \) is a noise term that represents a short-term deviation from the efficient price. Microstructure noise is usually an effect of market frictions, such as the difference between the bid and the ask prices (the bid-ask spread) and discretization errors. Various assumptions can be made on the serial correlation structure of \( \epsilon_t \), the most common of which is that \( \epsilon_t \) is a mean-zero white noise process. In this work, we will not concern ourselves much with specifics on \( \epsilon_t \) and instead will use a method described in Section 2.3 to minimize the effect of this noise on our measurements.
2.2 Estimators for the Variance

Various nonparametric estimators have been developed for the integrated variance and the variance of the jump process. The most common is the realized variance

\[ RV_t \equiv \sum_{j=1}^{M} r_{t,j}^2, \]  

where \( t \) indexes some period of time (such as days), \( r_{t,j} \equiv p_{t,j} - p_{t,j-1} \) is the geometric return between the \((j-1)\)st and \( j \)th price observations in a day, and \( M \) is the number of returns we observe in a single day. Barndorff-Nielsen and Shephard (2004) show that

\[ RV_t \overset{p}{\to} \int_{t-1}^{t} [\sigma(s)]^2 \, ds + \sum_{t-1<s\leq t} \kappa(s)^2. \]  

The first term in Equation 5 is called the integrated variance and denoted \( IV_t \). Thus, \( RV_t \) converges in probability to the integrated variance plus the sum of the squares of the jumps that occur within the day, which we call the total variance \( TV_t \).

Barndorff-Nielsen and Shephard also propose the bipower variation estimator

\[ BV_t \equiv \mu_k^{-2} \cdot \frac{M}{M-1} \cdot \sum_{j=2}^{M} |r_{t,j-1}| |r_{t,j}|, \]  

where \( \mu_k \equiv \mathbb{E}(|Z|^k) \) if \( Z \) is standard normal. Andersen, Dobrev, and Schaumburg (2010) propose the estimators

\[ MinVar_t \equiv \pi \cdot \frac{M}{M-1} \cdot \sum_{j=2}^{M} \min \{|r_{t,j-1}|, |r_{t,j}|\}^2 \]  

\[ MedVar_t \equiv \frac{\pi}{6 - 4\sqrt{3} + \pi} \cdot \frac{M}{M-2} \cdot \sum_{j=3}^{M} \med \{|r_{t,j-2}|, |r_{t,j-1}|, |r_{t,j}|\}^2. \]

It can be shown that all three estimators converge in probability to \( IV_t \). Therefore, these statistics can be used to define a measure of the relative contribution of jumps to the total price variance, \( IV_t/TV_t \). In
particular, we define

\[ \text{RJ}_{BV,t} \equiv \frac{RV_t - BV_t}{RV_t} \]  
(9)

\[ \text{RJ}_{Min,t} \equiv \frac{RV_t - \text{MinVar}_t}{RV_t} \]  
(10)

\[ \text{RJ}_{Med,t} \equiv \frac{RV_t - \text{MedVar}_t}{RV_t} \]  
(11)

to be three consistent estimators for this quantity. While all the above statistics are supposed to converge to a nonnegative quantity, they may be negative in finite samples. Andersen, Bollerslev, and Diebold (2007) truncate this quantity at 0 to account for this possible finite sample error. However, we will not perform this truncation in this paper, instead allowing for negative (unrealistic) values of \( \text{RJ}_{BV,t} \) if necessary. The rationale is that empirical investigations suggest that the statistic in which we are interested is not altered much by the truncation.

### 2.3 The Role of Microstructure Noise

Microstructure noise is an important part of much empirical work dealing with high-frequency data, and researchers have proposed many clever methods of dealing with this phenomenon. The method that we will employ in this paper was proposed by Andersen, Bollerslev, Diebold, and Labys (2000) and involves using data from \( N \)-minute intervals instead of the smallest 1-minute intervals available to us from the data. As we use finer sampling intervals, the microstructure noise (\( \epsilon_t \) in Equation 3) inflates the calculated variance. Under certain assumptions, it can even be shown that as the sampling interval approaches zero, the estimated variance will diverge. Therefore, plotting the unconditional realized variance as a function of the sampling interval usually exhibits the pattern that the variance increases for small (1-, 2-, or even 5-minute) intervals and approaches a flat, constant value for intervals that are sufficiently long. Such a volatility signature plot is useful for determining a sufficiently short interval that does not increase the unconditional volatility much above the value when using long intervals. Throughout this work, we will use 5-minute intervals.

Other methods have been developed to avoid having to throw away a portion \( (N - 1)/N \) of the data. However, since we do not use these estimators in this work, we will not discuss them in detail.

### 2.4 Jump Tests

Econometricians have proposed a large number of statistical tests to test the null hypothesis that the stock price in a particular trading day contains no jumps. Barndorff-Nielsen and Shephard (2004) propose the \( \text{RJ}_{BV,t} \) statistic as part of a jump test; Huang and Tauchen (2005) modify this test statistic slightly and
propose

\[
Z = \frac{R_{J_{BV,t}}}{\sqrt{\left(\frac{\pi^2}{4} + \pi + 5\right) \cdot \frac{1}{n} \cdot \max\left\{1, \hat{Q}_{BV_t}\right\}}}.
\]  

(12)

where \( \hat{Q} \) is a consistent estimator for the integrated quarticity \( \int_{t-1}^{t} [\sigma(s)]^4 ds \). The two estimators Tauchen and Huang consider are the tripower quarticity and the quadpower quarticity, given by

\[
\hat{Q}_k = M \cdot \mu_{4/k} \cdot \frac{M}{M - k + 1} \sum_{i=k+1}^{M} \prod_{j=1}^{k} |r_{i-j}|^{4/k},
\]  

(13)

for \( k = 3 \) and \( k = 4 \), respectively. The test statistic is asymptotically standard normal under the null hypothesis. Huang and Tauchen’s Monte Carlo simulations support the use of tripower quarticity in finite samples.

Since we are interested in studying properties of \( R_{J_{BV,t}} \) and will eventually compare them to results from jump tests, the B-NS test poses an issue, since any artifact that may affect \( R_{J_{BV,t}} \) is finite samples will also affect the outcome of the jump test directly. Instead, we will use a test proposed by Jiang and Oomen (2008) that exploits higher moments of the return process. If a process is continuous, these higher moments should be small whereas in the presence of jumps, these moments will be nontrivial. The statistic they consider is called \textit{swap variance} and is calculated as

\[
SwV_t = 2 \cdot \sum_{i=2}^{M} (R_{t,i} - r_{t,i}),
\]  

(14)

where \( R_{t,i} \) is the arithmetic return from time \( i - 1 \) to time \( i \) within a day \( t \). Thus,

\[
R_{t,i} = \frac{P_{t,i} - P_{t,i-1}}{P_{t,i-1}} = \frac{P_{t,i}}{P_{t,i-1}} - 1 = \exp (p_{t,i} - p_{t,i-1}) - 1 = \exp (r_{t,i}) - 1.
\]

A Taylor expansion of the exponential function shows that \( SwV - RV \) contains only terms of order \( r_{t,i}^3 \) and higher. Thus, we would expect \( SwV \approx RV \) in the absence of jumps. From this observation, we can studentize this distribution and obtain the statistic

\[
JO = M \cdot \frac{BV_t}{\sqrt{\hat{\Omega}_{SwV}}} \left(1 - \frac{RV_t}{SwV_t}\right),
\]  

(15)

where \( \hat{\Omega} \) is a consistent estimator for the integrated sexticity \( \int_{t-1}^{t} [\sigma(s)]^6 ds \). Possible estimators for this
quantity include
\[
\hat{q}^{(p)}_{SwV} = M^2 \cdot \frac{\mu_{6}}{9} \cdot \mu_{p-6} \cdot \frac{M^2}{M - p + 1} \sum_{j=p+1}^{M} \prod_{k=1}^{p} |r_{t,j-k}|^{6/p},
\]

with \( p = 4 \) and 6 being the obvious choices. We will use \( p = 4 \) in this work. The statistic \( JO \) is asymptotically standard normal.

Jiang and Oomen propose other test statistics as well, all of which exploit the observation that \( SwV \approx RV \) under the null. Properly scaled, these statistics are all asymptotically standard normal. Monte Carlo simulations conducted by Jiang and Oomen using a stochastic variance model of a mean-reverting square-root process support the statistic given in Equation 15 due to its finite sample properties.

3 Initial Data Analysis

3.1 Data

We have access to minute-by-minute data for 22 stocks in the S&P 500. We choose these stocks in three major industry sectors—technology, finance, and food/agricultural—although we do not use any quantitative criteria to choose them. The price data for each stock is obtained from a commercial data vendor. Each trading day contains price data from 9:35 AM to 4:00 PM; while trading does start at 9:30 AM each day, we ignore the initial five minutes as market reaction to overnight events may be substantially different from its intraday behavior. Most stock data runs from 1997 to 2010, although certain securities are available for only a subset of that time period. For example, GOOG is obviously available only after its IPO in 2004, and XOM is available only after the Exxon/Mobil merger in 1999.

The stock data has been adjusted backward for stock splits, using dates from Yahoo! Finance as a guide to distinguish stock splits from large jumps and other economic events. Days with data entry errors (missing price values, for example) are very rare but are nonetheless excluded from the analysis. A number of trading days are missing for certain stocks. As a result, when calculating statistics that rely on data from two stocks, care is taken to align the time series properly: any date that is missing is one series is discarded from the other series as well.

Table 1 lists the stocks we choose along with their industries, start dates, end dates, and number of trading days observed.

\[\text{A stock split is also distinguishable in the data since the stock price falls by exactly one-half or two-thirds.}\]
3.2 Univariate $RJ_t$ Series

We begin with a brief analysis of time series of $RJ_{BV,t}$ for individual stocks. Figure 1 plots the time series of $RJ_{BV,t}$ for KO. The time series is at the daily level, and the realized and bipower variation were calculated using 5-minute intervals. This plot is representative of the time series for most stocks in that it seems somewhat similar across time. The mean of the series is positive even though the statistic drops below zero for many days (27.4% of days have $RJ_{BV,t} < 0$). Table 2 tabulates the mean of the $RJ_{BV}$ series for various stocks by year, from 1997 to 2010. This table also calculates the means and standard errors across all stocks in each of the three industries. We note a few general observations:

- The average value of $RJ_{BV,t}$ from 1997–2010 is approximately equal across industries. In fact, given the variation in the statistic within industries, the difference in the industry mean of $RJ_{BV,t}$ is likely statistically insignificant.\(^2\)

- Some stocks do show a noticeable and significant time trend. For example, AAPL, BK, and HPQ—among others—have mean values of $RJ_{BV}$ that are higher in the earlier portion of the sample (1997 to about 2001) than in the latter portion of the sample. Indeed, the difference between BK’s mean in 1997 and 2006 is certainly statistically significant at the 5% level. Furthermore, a similar but less significant trend can be seen in most of the stocks: the average value for $RJ_{BV}$ seems to decrease for most stocks around 2001–2003, and it stays roughly constant from then on. This trend is noticeable when averaging across stocks as well, as can be seen from final four lines of the table, although large within-industry variations induce larger standard errors in these averaged quantities.

The data in Table 2 suggest that the statement that $RJ_t$ is similar across stocks and across time is not entirely accurate, although the mean value of $RJ_t$ seems to follow a similar pattern for most stocks. While we do not show the data in this paper, similar trends happen with $RJ_{Min,t}$ and $RJ_{Med,t}$. However, the time trend is not as prominent when considering $RJ_{Min,t}$.

Figure 1 also plots kernel density estimates for the distribution of $RJ_{BV,t}$, $RJ_{Min,t}$, and $RJ_{Med,t}$ calculated from the data for KO. This plot is also representative of those of most other stocks: we observe that all three distributions have similar means but that the distribution for $RJ_{BV}$ is much narrower than that for $RJ_{MedV}$ and $RJ_{MinV}$. Furthermore, an interesting observation is that a Kolmogorov-Smirnov test rejects the null hypothesis that the $RJ_{BV,t}$ statistic is distributed normally at the 5% level for all stocks; however, testing each year separately does not reject the null hypothesis for any stock during any year. Performing the same

\(^2\)A two sample $t$-test would return that the difference between the “overall” values for the food and finance industries is insignificant at the 5% level. However, we need the assumption that the samples are independent in order to run a naive two-sample $t$-test. I do not believe this assumption should hold in our case. A more formal hypothesis test would be to construct the time series $D_t = R_{food} - R_{fin}$, where the average is over different stocks at a particular day, and test whether the mean of this time series is 0 using standard techniques.
K-S test on the $RJ_{Min,t}$ statistic does not reject the null hypothesis. We conjecture that since the time trend in a particular stock is more noticeable when using the $RJ_{BV,t}$ statistic, the values for each individual year of a particular stock may be distributed normally with significantly different means. Collecting these values together therefore results in a single distribution that is significantly not normal.

Finally, we compute autocorrelation functions for the $RJ_{BV,t}$, $RJ_{Med,t}$, and $RJ_{Min,t}$ time series for each stock. This investigation is motivated by a study conducted by Bollerslev, Kretschmer, Pigorsch, and Tauchen (2009) in which they calculate the autocorrelation function for the statistic $J_t \equiv \log(RV_t/BV_t)$ using the S&P futures data. They find that fifth-order autocorrelations are significant. An explanation for this observation is that economic data is often released in a weekly fashion, and this weekly periodicity corresponds to the fifth-order lag. We thus apply the same analysis to $RJ_t$ for individual stocks. We find that all autocorrelations above zeroth order are insignificant when using the simple Bartlett confidence intervals. Owing to this uninteresting result, we do not show an example plot of a typical autocorrelation function.

### 3.3 Bivariate $RJ_t$ Series

A natural question to consider is whether $RJ_t$ series for different stocks are related; this is the question that we will pursue in this paper. Stanley (2007), in a thesis written in this class, briefly considers a somewhat related question—albeit in a different context. He plots the $z$-statistic (Equation 12) for a number of common stocks against that for the S&P 500 and concludes that there is no relation. However, the S&P 500 represents an aggregate index, and a jump in an aggregate index is usually a result of large cojumps in a single stock or a number of smaller cojumps in a larger number of stocks. The question that we will explore is whether there is a correlation between $RJ_t$ series in two different stocks.

For convenience, for two stocks with $RJ_{BV,t}$ series denoted $RJ_{(1)BV,t}$ and $RJ_{(2)BV,t}$, define

$$C_{BV} \equiv \text{corr} \left( R_{(1)BV,t}, R_{(2)BV,t} \right)$$

(17)

to be the zeroth order cross-correlation of the two time series. We will define $C_{MinV}$ and $C_{MedV}$ similarly. Furthermore, we could consider $\tau_{BV}$ and $\rho_{BV}$ as alternate estimators for this quantity, where $\tau$ and $\rho$ denote correlation calculated via Kendall’s $\tau$ and Spearman’s $\rho$ instead of Pearson’s coefficient. Instead of calculating the $RJ_t$ series over days, we can calculate it over weeks or months; to be explicit, this new time series will be an estimate for the portion of weekly (or monthly) volatility that is due to jumps. These alternate time series

---

\(^3\)Note that a first-order expansion of $J_t$ around 1 shows that $-J_t \approx \frac{BV_t}{RV_t} - 1 = -RJ_t$.  

\(^4\)These alternate nonparametric estimators of correlation were discussed in detail during the presentations during the course of the semester. To save space and keep this report terse, we will simply summarize our observations using these estimators here and in Section 5 as a side note.
can be used to calculate $C_{BV}$, $C_{Min}$, and $C_{Med}$.

To calculate $C_{BV}^{(d)}$ for pairs of stocks, we first align the time series. Any day that is not present in either stock is discarded from the data, and the correlation is calculated using only the days present in both stocks. Weekly correlations $C_{BV}^{(w)}$ are calculated only using weeks when five consecutive days (Monday through Friday) are present in both stocks. Finally, monthly correlations $C_{BV}^{(m)}$ are calculated using groups of four consecutive full weeks present in both stocks.

A graphical description of some results is given in Figure 2. This figure plots $C_{BV}^{(d)}$, $C_{BV}^{(w)}$, and $C_{BV}^{(m)}$ simultaneously and color-codes the industries, as described in the caption. The stocks BAC (a finance company) and DELL (a tech company) were chosen since they exemplify two different patterns. Considering the plot for BAC, we see that finance stocks tend to have a higher $C_{BV}$ with BAC than many other stocks, as these points are clustered closer to the right-hand side of the figure. Furthermore, there seems to be a clear pattern as we move from $C_{BV}^{(d)}$ to $C_{BV}^{(w)}$ to $C_{BV}^{(m)}$ in that the correlation grows stronger. In the plot for DELL, the temporal pattern is still qualitatively evident, but it is less clear whether tech stocks have a higher $C_{BV}$ with DELL.

To quantify this relation, for each stock, we calculate the average $C_{BV}$ by industry. The results of this calculation are given in Table 3. While we have not summarized the results from formal statistical tests of the hypothesis that this average quantity is larger within-industry than out-of-industry, visual inspection of the means and the standard errors can offer some insight. We see that the observation that within-industry correlations are larger seems to be metted out in a number of stocks, especially food and finance ones. However, in many of these cases, the difference is unlikely to be statistically significant if using a simple two-sample $t$-test. We must note that with as few as 5–6 datapoints per industry, one errant observation (misclassifying T as the same type of “tech” stock as GOOG or AAPL) can alter the mean greatly and increase the standard error. As a result, we will take this observation as simply a guiding principle that “similar” stocks tend to have higher $C_{BV}$ values, although all the comparisons in this paper will compare within-industry to out-of-industry. Finally, the reason we list all stocks in Table 3 instead of averaging within industries is that certain stocks seem to have systematically higher $C_{BV}$ values than other stocks—both within- and out-of-industry. An additional average, while more compact, would potentially confound the results with the effects of such level differences.

---

5 We will often denote by $C_{BV}^{(d)}$ the correlation calculated at the daily level, by $C_{BV}^{(w)}$ that calculated at the weekly level, and $C_{BV}^{(m)}$ that at the monthly level. However, when the time period is clear from context—or if it is arbitrary—we will omit the superscript to limit the number of adornments to our variables.

6 We have neglected standard error on the correlations. As an estimate, we can consider the Bartlett rule and assign errors of $2/\sqrt{N}$ to form a 95% confidence interval, if $N$ is the number of observations. Note that $N \sim 2500$ on the daily level, $N \sim 500$ for the weekly level, and $N \sim 120$ for the monthly level. As a result, the confidence interval at the monthly level is quite large. However, the fact that this pattern is replicated in many different stocks encourages us to think about this pattern in the statistic despite the large errors.
One method to formally test the claim that $C_{BV}^{(d)} < C_{BV}^{(w)}$ is to run the regression $C_{BV}^{(w)} = \alpha + \beta \cdot C_{BV}^{(d)}$ where the datapoints are all $\binom{22}{2} = 231$ pairs of stocks. We compare the null hypothesis $H_0 : \beta \leq 1$ against the alternative $H_1 : \beta > 1$. The scatters for $C_{BV}^{(w)}$ vs. $C_{BV}^{(d)}$ and $C_{BV}^{(m)}$ vs. $C_{BV}^{(d)}$ are shown in Figure 3. At any reasonable significance level, we can reject the null hypothesis.

A possible hypothesis for these observations is that the higher correlation is due to jumps that either occur simultaneously or during the same day. We can compare the correlations we found above with correlations between two time series of indicator variables. For a stock labeled $i$, let $\text{Jump}_{i}^{(1)}$ be the event that stock $i$ has a jump in day $t$. We can calculate the indicators $\mathbb{1}[\text{Jump}_{i}^{(1)}]$, using a Jiang-Oomen test at a 1% or a 0.1% significance level. Then, we can compare $JC \equiv \text{corr}\left(\mathbb{1}[\text{Jump}_{i}^{(1)}], \mathbb{1}[\text{Jump}_{i}^{(2)}]\right)$ to $C_{\ast}^{(d)}$. A positive association between these statistics would suggest that $C_{\ast}$ may be related to the lack of independence between the underlying jump processes.\footnote{For the sake of completeness, we emphasize that in order for the correlation $JC$ to be large, it is not necessary for jumps to occur simultaneously. Even if the two stocks are likely to have jumps that occur during the same day, if not simultaneously, the “jump” correlation will be large. During the course of the semester, more tests were done using statistical methods to test for common arrival of jumps, such as the one proposed by Jacod and Todorov (2009). These results are not presented in the paper since they were difficult to interpret, and no clear pattern was found.}

Figure 4 plots example comparisons between $C_{MinV}^{(d)}$ and $JC$ at the 1% and 0.1% significance levels. While the figure for BAC shows a similar industry ordering between $JC$ and $C_{MinV}^{(d)}$, closer inspection suggests that the within-industry ordering differs between the two statistics. The data for MON shows a starker difference between $JC$ and $C_{MinV}^{(d)}$. To summarize these comparisons, Figure 4 also includes a scatterplot of $JC$, calculated at the 1% significance level, against $\tau_{BV}^{(d)}$ for all 231 pairs of stocks. The slope of a regression line through this scatter—as can be anticipated from the diagram—is insignificant at any reasonable significance level. Indeed, replacing $\tau_{BV}^{(d)}$ with any of the other eight statistics we can consider at the daily level does not yield a significant linear relation between the two statistics. We conclude that the statistic $C_{BV}$ (and all related statistics) contains information that cannot be explained purely by tests for the presence of jumps. We conjecture that quantities such as size and correlation in magnitudes may also be important.

As a final note for this data analysis, we mentioned earlier that we could also consider $C_{MinV}$ and $C_{MedV}$, along with $\tau_{BV}$ and $\rho_{BV}$. In the previous paragraph, we have utilized these estimators simply to offer an alternative to $C_{BV}$. The general result from consider these statistics is that using $MinVar$ and $MedVar$ depress the correlation slightly as compared to $C_{BV}$. A similar observation is seen when using the nonparametric measures of correlation. However, the observations that we have discussed still hold.

In summary, the work this semester has been motivated by three main observations:

- $C_{\ast}$ tends to be higher within industry than out-of-industry, although in our small sample, the difference for many stocks is not statistically significant.
• It is usually the case that $C_{BV}^{(d)} < C_{BV}^{(w)} < C_{BV}^{(m)}$.

• The alternate statistic $JC \equiv \text{corr} \left( \mathbb{1} \left[ \text{Jump}_t^{(1)} \right], \mathbb{1} \left[ \text{Jump}_t^{(2)} \right] \right)$ does not exhibit any significant correlation with $C_{BV}^{(d)}$, which suggests that $C_{BV}$ is significantly affected by some effect other than simply the presence of jumps.

The final observation is not unanticipated, given that $RJ_{BV,i}$ is designed to measure the magnitude of the jump variation with respect to the total variation. These qualitative observations motivate our Monte Carlo studies in the next section.

4 Monte Carlo Simulations

We can try to explain these observations in a simple framework in the absence of jumps. Suppose we have two stocks, labeled $i = 1, 2$. Assume that the return process both both stocks has no underlying drift and that $\sigma^{(1)}(t)$ is deterministic and constant at $\sigma^{(1)}$. Furthermore, assume that the underlying Brownian motion processes have correlation $\rho$ so that $r_j^{(i)} = \sigma^{(i)} Z_j^{(i)}$ where

$$
\begin{pmatrix}
Z_j^{(1)} \\
Z_j^{(2)}
\end{pmatrix}
\sim \mathcal{N}
\left(
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix}
\right),
$$

and the vector $(Z_j^{(1)}, Z_j^{(2)})'$ is iid. Since $C_{BV}$ can be expressed as $\text{corr} \left( BV_t^{(1)} / RV_t^{(1)}, BV_t^{(2)} / RV_t^{(2)} \right)$, we are interested in the expressions

$$
\mathbb{E} \left[ \frac{\mu_1^{-2} \sum_{j=2}^M \left| r_{t,j}^{(1)} \right|^2 \left| r_{t,j-1}^{(2)} \right|^2}{\sum_{j=1}^M \left( r_{t,j}^{(1)} \right)^2} \right] = \mu_1^{-4} \mathbb{E} \left[ \frac{\sum_{j=2}^M \left| Z_{t,j}^{(1)} \right|^2 \left| Z_{t,j-1}^{(2)} \right|^2}{\sum_{j=1}^M \left( Z_{t,j}^{(1)} \right)^2} \right] \quad (18)
$$

and

$$
\mathbb{E} \left[ \left( \frac{\mu_1^{-2} \sum_{j=2}^M \left| Z_{t,j}^{(1)} \right|^2 \left| Z_{t,j-1}^{(2)} \right|^2}{\sum_{j=1}^M \left( Z_{t,j}^{(1)} \right)^2} \right)^2 \right] = \mu_1^{-4} \mathbb{E} \left[ \left( \frac{\sum_{j=2}^M \left| Z_{t,j}^{(1)} \right|^2 \left| Z_{t,j-1}^{(2)} \right|^2}{\sum_{j=1}^M \left( Z_{t,j}^{(1)} \right)^2} \right)^2 \right]. \quad (19)
$$

It is clear that $C_{BV}$ can be expressed in terms of Equations 18 and 19. Unfortunately, both these equations are very complicated and difficult to simplify analytically because of the ratio. Consider simply Equation 19. While the denominator has a $\chi^2(M - 1)$ distribution and it may be possible to derive the distribution of the numerator, the numerator and denominator are not independent. Any analytic progress on this problem

---

8The work that we show in this brief introduction can easily be changed to account for a time-dependent deterministic volatility. We can write the required expectations purely in terms of standard normals in the case of constant volatility, while in the case of nonconstant volatility, we will not have complete cancellation of the $\sigma(t)$ series.
seems especially difficult and possibly intractable.

We resort to Monte Carlo simulations to study these statistics in more detail. The following subsections discuss the simulations that we conduct.

### 4.1 Setup

We use a simple Monte Carlo model to study the dependence of $C_{BV}$ on various parameters. Consider two stocks labeled 1 and 2. We first assume that the underlying Brownian motions are correlated with correlation $\rho_d$. That is, say \((Z_t^{(1)}, Z_t^{(2)})')\ is iid with distribution

$$\begin{pmatrix} Z_t^{(1)} \\ Z_t^{(2)} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_d \\ \rho_d & 1 \end{pmatrix} \right). \quad (20)$$

We have three jump processes that are mutually independent. Let the idiosyncratic jump processes be denoted $dJ_t^{(i)} = \kappa_i(t) dq_i(t)$ for $i \in \{1, 2\}$, where the jump magnitude $\kappa_i(t) \sim N(0, \sigma_{jump,i}^2)$ and $q_i(t)$ is a Poisson counting process with parameter $\lambda_i$. We also have a common jump component $dCJ_t = (dCJ_t^{(1)}, dCJ_t^{(2)})' = \kappa_c(t) dq_c(t)$ where $q_c(t)$ is a Poisson counting process with parameter $\lambda_c$ and

$$\kappa_c(t) \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{cj,1}^2 & \rho_c \sigma_{cj,1} \sigma_{cj,2} \\ \rho_c \sigma_{cj,1} \sigma_{cj,2} & \sigma_{cj,2}^2 \end{pmatrix} \right). \quad (21)$$

Of course, all the jump magnitudes ($\kappa_1, \kappa_2$, and $\kappa_c$) are mutually independent, and all counting processes are also mutually independent.

The price is then simulated as

$$\Delta p_t^{(i)} = \sigma_i Z_t^{(i)} + dJ_t^{(i)} + dCJ_t^{(i)}, \quad (22)$$

where $\sigma_i$ is the volatility of the diffusive component of stock $i$. In each of these simulations, we generate 1000 sample paths for $252 \times 5$ trading days. We record prices at 5-minute increments to emulate our treatment of the data.

A summary of the notation in the setup is given in Table 4. In later sections, we will discuss drawbacks and limitations of this simulation setup.
4.2 Parameter Studies

We first study the dependence of various parameters on $C_{BV}$, $C_{Min}$, and $C_{Med}$. In these cases, for simplicity, we let $\sigma_1 = \sigma_2$, $\lambda_1 = \lambda_2$, and $\sigma_{jump,1} = \sigma_{jump,2} = \sigma_{cj,1} = \sigma_{cj,2}$. The parameters we use are given in Table 5.

4.3 Direct Comparisons to Data

While the setup in Section 4.2 is designed to offer insight into the effects of parameters, it does not allow for direct comparison to the data. That is, it would be interesting to determine how this Monte Carlo simulation, with parameters calibrated from the data, compares to the observed values of $C_{BV}^{(d)}$, $C_{BV}^{(w)}$, and $C_{BV}^{(m)}$. The expectation is not perfect numerical agreement with the data. Rather, systematic deviations from the data may suggest directions for future investigation.

To run these comparisons, we must estimate values for the parameters introduced in Section 4.1 from the data. To estimate $\sigma_i$, we calculate the bipower volatility of the individual stock. The diffusive covariance $\rho_d$ is estimated through bipower covariance as calculated through the portfolio method. That is, if $r_A$ and $r_B$ are five-minute returns of stocks $A$ and $B$, we create a balanced “portfolio” of those stocks such that the five-minute returns are given by $\frac{1}{2} r_A + \frac{1}{2} r_B$. Noting that

$$\var\left(\frac{1}{2} r_A + \frac{1}{2} r_B\right) = \frac{1}{4} \var(r_A) + \frac{1}{4} \var(r_B) + \frac{1}{2} \cov(r_A, r_B),$$

we can back out the bipower covariance (the final term) by using bipower variance to calculate the first three terms. Then, our estimate for $\rho_d$ is simply $\hat{\rho}_d = \cov(r_A, r_B) / \left[\var(r_A) \cdot \var(r_B)\right]^{1/2}$. The frequency parameters $\lambda_i$ and $\lambda_c$ are estimated using a Jiang-Oomen test at a 0.5% significance level. We say that $\hat{\lambda}_i$ is the number of “jump days” identified in a particular stock divided by the total number of days. Similarly, $\hat{\lambda}_c$ is the number of days during which both stocks jump divided by the total number of days on which both stocks are observed.

The remaining parameters require identification of the magnitudes of observed jumps. To identify jumps, we take a very naive approach proposed by Andersen, Bollerslev, and Dobrev (2006). Under the simplifying assumption that volatility is constant within the trading day, they show that a randomly drawn intraday diffusive return is approximately distributed $N(0, \Delta \cdot BV_t)$, where $\Delta$ is the sampling frequency. In the case of 5-minute intervals—as we use in this paper—we set $\Delta = 1/77$. Thus, to detect intraday jumps, Andersen, Bollerslev, and Dobrev propose

$$\kappa_j = r_{j,t} \cdot I[|r_{j,t}| > \Phi_{1-\beta/2} \cdot \sqrt{\Delta \cdot BV_t}],$$

where $I[.]$ is an indicator function that takes the value 1 if the condition inside the brackets is true, and 0 otherwise.
where $\beta$ is the significance level of the test. As suggested by Andersen, Bollerslev, and Dobrev, we use $\beta = 1 - (1 - \alpha)^{\Delta}$ with $\alpha = 10^{-5}$. Intuitively, Equation 24 identifies as jumps any movements that are sufficiently large so as to be unlikely to be due to the diffusive process.

Once these jumps are identified, estimating $\sigma_{\text{jump},i}^2$, $\sigma_{c,i}^2$, and $\rho_j$ is straightforward. Cojumps are identified as jumps that occur during the same 5-minute interval. The sample correlation between the magnitude of these cojumps is our estimate for $\rho_j$. The same standard deviation of the magnitude of the jumps in stock $i$ identified as “cojumps” is our estimate for $\sigma_{c,i}$. The sample standard deviation for any jumps in stock $i$ not identified as cojumps is our estimate for $\sigma_{\text{jump},i}$. This method of identifying jumps also provides a manner of cross-checking our estimates for $\lambda_i$ and $\lambda_c$, since we can compare the number of “jump” intervals to the total number of intervals. In general, the two estimation procedures are close to each other, so we adhere to the one previously outlined using the Jiang-Oomen test for our estimates in this work.

Table 4 summarizes the estimation methodology as well. Appendix A discusses an issue with the estimation procedure and model for jumps: specifically, modeling the jumps as normal instead of truncated normal introduces some technical problems. However, we adhere to the model in Section 4.1 simply because it may serve as an easy-to-interpret first step towards understanding this correlation. Furthermore, it prevents us from having to interpret yet another parameter in the truncation level of the jumps.

5 Results

We separate the results into those obtained from the setup in Section 4.2 and those from the setup in Section 4.3.

5.1 Parameter Studies

Given that in our Monte Carlo simulations, we leave only four adjustable parameters—$\sigma_c$ (referred to as $\sigma_{c,i}$ in this section), $\lambda_c$, $\rho_d$, and $\rho_c$—we will present results for the parameter investigations in Section 4.2 in graphical format, plotting $C_{BV}^{(d)}$ against either $\rho_d$ or $\rho_c$ for various values of $\sigma_c$. The goal of this analysis is to gain intuition about how these parameters can affect $C_{BV}$.

Figure 5 summarizes the effect of both $\rho_d$ and $\lambda_c$ on $C_{BV}$. We fix $\rho_c = 0.5$ and show $C_{BV}$ against $\rho_d$ for various values of $\sigma_c$. The first somewhat surprising observation is that $C_{BV}$ has a strong positive dependence on $\rho_d$. That is, even when jumps are very small—volatility in a single jump is less than the daily volatility in the stock—, a large correlation in the underlying diffusive component can induce a residual correlation

\footnote{Note that since we impose that the distribution of the jumps has mean 0, we actually use the sample second moment as our estimators for the standard deviation parameters. However, the difference between the sample standard moments and the sample standard deviations is small since the sample means are close to zero and statistically insignificant.}
between $RJ_{BV,t}$ series and increase $C_{BV}$ to a statistically significant positive number.\textsuperscript{10} Note, as will be discussed in the subsequent section, that most pairs of stocks have bipower correlation around 0.2–0.5, which is within the range of the parameter space where $C_{BV}$ is noticeably positive for even the lowest value of $\lambda_c$ considered. The most reasonable explanation for this observation is that the statistic $RJ_{BV,t}$ picks up large diffusive movements as “jump” movements during the calculation. If the diffusive motion is correlated, stocks are more likely to have these large diffusive movements in the same day.\textsuperscript{11}

Another pattern evident from these simulations is that an increase in $\sigma_c$ increases $C_{BV}$. The intuition for this result is clear: in the case that there is a jump within a day, the contribution to the total variation by this jump will be larger if $\sigma_c$ is larger. Potentially since $\sigma_c$ dials the magnitude of the cojumps as well, both $RJ_{BV,t}^{(1)}$ and $RJ_{BV,t}^{(2)}$ will be large on this day. Introducing ordered pairs where both elements are large will increase the correlation between the time series. Finally, we note that an increase in $\lambda_c$ increases $C_{BV}$. The argument given above for increases in $\sigma_c$ applies in this case as well. If common jumps are more frequent, then a greater number of days will have large values of both $RJ_{BV,t}^{(1)}$ and $RJ_{BV,t}^{(2)}$. The effect of changing $\lambda_c$ from 0.075 (one jump about every 13 days) to 0.25 (one jump every four days) to 1 is significant. For the sake of comparison, the value of $\lambda_c$ that we see in the data is on the lower end of this range.

Another dimension we can consider is altering $\rho_c$, which had been fixed at 0.50 in our previous analysis. Figure 6 shows similar figures with $\rho_d$ fixed at 0.40, a value that is comparable to that for two stocks in the same industry. Here, we again see the effect of higher $\sigma_c$ and $\lambda_c$, and the explanations are similar to those given in the previous paragraphs. The interesting observation in this parameter study is that the effect of $\rho_c$ on $C_{BV}$ seems to be minimal; that is, having cojumps with highly correlated magnitudes does not increase $C_{BV}$ to nearly the same extent that increasing $\rho_d$ does. A potential explanation of this unexpected observation is that introducing cojumps of sufficiently large magnitudes already increases $RJ_{BV,t}^{(i)}$ for both $i = 1$ and 2 beyond days without cojumps. An increase in the correlation $\rho_c$ will thus be most significant when $\sigma_c$ is small, since in these cases, the event that a cojump consists of one large jump and one “jump” only slightly larger than the diffusive variation is more likely. We do observe this dependency in the simulations, as the curve for small $\sigma_c$ slopes upward as $\rho_c$ increases above 0.6. However, some simple mathematical model testing this conjectured explanation would be useful.

We can now make a few comments about the observations presented at the end of Section 3.3:

- The within-industry pattern may be due to a larger frequency of cojumps, but it may also be something

\textsuperscript{10}To simplify the discussion, we do not explicitly worry about error bars in our discussion, although it is a very important component of statistical analysis. We can estimate an error on the simulated values of $C_{BV}$ by looking at the distribution of the results from the 1000 sample paths. The empirical distribution of this statistic over these 1000 replicates usually has a standard deviation around 0.05. Scaling by $1/\sqrt{1000}$, the increase we see over $\rho_d$ would be clearly significant at any reasonable level.

\textsuperscript{11}A natural conjecture is that this increase may purely be due to poor performance by BV in $RJ_{BV,t}$ in identifying the diffusive component of the variation. However, $C_{MinV}$ and $C_{MedV}$ show similar patterns.
as simple as an artifact of a large correlation in the diffusive movements in stocks in the same industry.

- We have not discussed quantities such as $C_{BV}^{(w)}$ and $C_{BV}^{(m)}$ in this section. The general result is that it is possible to replicate the increase from $C_{BV}^{(d)}$ to $C_{BV}^{(m)}$ with proper choice of parameters, but it is also possible to induce the opposite observation by carefully choosing a different parameter set.

- The statistic $JC$ presented in Section 3.3 only tests for the presence of jumps—not their size or the correlation in the magnitude of cojumps. We see that $C_{BV}$ is strongly affected by the size of the jumps in the data series, as we would expect. However, it seems that the dependence on the correlation in cojumps is minimal and almost nonexistent for sufficiently large $\sigma_e$.

### 5.2 Comparisons to Data

A large simplification in the model studied in the previous section is that we imposed constraints on our parameters to limit the parameter space to a four-dimensional region. In this section, we move from this 4-dimensional parameter space to a larger 11-dimensional space over which all the parameters introduced in Section 4.1 are allowed to vary. A consequence of this large parameter space is that it is difficult to analyze the effect of all these parameters by performing work similar to what we did in Section 5.1.\(^{12}\) Thus, we will compare simulation moments to moments in the data and identify systematic deviations.

We begin with a brief discussion of general results from this setup: we generate graphs similar to those in Figure 2 from the simulated runs instead of the observed data. Figure 7 shows these results for the same stocks (BAC and DELL). A few observations are clear: with the parameters for BAC, we see neither the dispersion in $C_{BV}$ evident in the data nor the increase in $C_{BV}$ from days to weeks to months also evident in the data. It is possible to assign standard errors to the means plotted in Figure 7, since we can obtain a distribution of $C_{BV}$ from the 1000 different sample paths. Regardless of the pair of stocks considered, the standard deviation of the distribution of $C_{BV}^{(d)}$ is very close to 0.028, that of $C_{BV}^{(w)}$ is close to 0.064, and that for $C_{BV}^{(m)}$ is around 0.14. Scaling these by $1/\sqrt{1000}$ to obtain standard errors of the mean, we see that the standard error assigned to the points at the daily level is very small; indeed, a hypothesis test comparing $C_{BV}^{(d)}$ of BAC with JPM to that for BAC with IBM (the highest tech stock in the diagram) would reject the null of equality at any reasonable significance level. However, it is unlikely that the difference we observe is economically significant. On the other hand, the simulations for DELL show a clear industry ordering in that many tech stocks have a (statistically significantly) larger $C_{BV}$ than most other stocks at daily, weekly, and

---

\(^{12}\)One clear approach is to reduce the dimension of this parameter space by looking at the dependence of $C_{BV}$ on economically meaningful combinations of the individual parameters. An example might be $P \equiv \lambda_c \sigma_{c,1}^2 / \left( \lambda_c \sigma_{c,1}^2 + \sigma_1^2 + \lambda_1 \sigma_{jump,1}^2 \right)$, which is the proportion of the variance of stock 1 in a particular day attributable to the cojump. However, designing these quantities requires time, and I have not made much progress on this endeavor this semester. The few I have considered have not yielded any interesting results.
monthly levels. Furthermore, the increase in the magnitude of this statistic from daily to monthly levels is much more significant than in the BAC case. Indeed, we see that the simulated $C_{BV}$ actually overshoots the $C_{BV}$ calculated from the observed data.

It is difficult to do a thorough comparison of $C_{BV}$ of all stocks when correlated with DELL, for example, so we will choose two pairs: BAC/GS and DELL/IBM. The pair BAC/GS has a statistically significant $C_{BV}^{(d)}$ in the observed data, and it increases significantly when moving to weekly and monthly periods. However, we do not observe this increase at all in the simulated moments, and the statistic remains below 0.1 in the simulations at all time levels. On the other hand, $C_{BV}$ for DELL/IBM significantly increases in the simulations for longer time periods, while it remains at low levels and shows an insignificant trend in the observed data. Table 6 lists the parameters for these two pairs of stocks. We see that both pairs have very similar values of $\rho_d$ and $\rho_c$ and also have comparable values of the diffusive volatility. The clearest systematic difference comes from comparing $\sigma_{c,j}$ for the DELL/IBM pair to that for the BAC/GS pair. The volatility of the cojumps is almost twice as large in the case of DELL/IBM. Indeed, this statistic is large when considering DELL with most pairs of tech stocks. That $\sigma_{c,j}$ affects $C_{BV}$ is understandable from our discussion from Section 5.1. It would be fruitful to do a similar detailed comparison of other pairs of stocks, but we have not conducted such analysis this semester.

Instead, we will present a few sample comparisons with patterns representative of most other pairs. Table 7 shows results for a comparison of BAC and C, both finance stocks. We list the mean and standard deviation of the distributions of $C_{BV}$, $C_{MinV}$, and $C_{MedV}$. It is useful to think of the observed data as one instance of this sample; that is, under the hypothesis that the data can be modeled using the setup in Section 4.1, the true data is simply one sample path. Comparing simulated results with the observed data in Table 7 shows a discrepancy that is quite common: the statistics from the observed data are larger than those from the simulations. Furthermore, this difference grows as we move from daily to monthly periods, usually even when normalizing the difference by the reported standard deviation. This “undershoot” error is the most common type of discrepancy we see. As a different type of example, Table 8 shows the same data for the pair BK/KFT, which are stocks from different industries. Here, the agreement is quite close: the observed data points lie within one standard deviation of the simulated mean for all the listed statistics. Perhaps more importantly, there does not seem to be a systematic sign to the quantity $C_{sim}^* - C_{data}^*$, meaning the simulations tend to neither “overshoot” nor “undershoot” the data. Indeed, this pattern is common in many out-of-industry comparisons.

We attempt to quantify the observation motivated in the previous paragraph by considering the quantity

\footnote{Again, we can use the Bartlett confidence interval at a 5\% level, for concreteness.}
\footnote{The opposite “overshoot” error does also occur, but it is much less frequent. The DELL/IBM correlation mentioned above is one of the few examples.}
\[|C_{BV}^{\text{sim}} - C_{BV}^{\text{data}}|\] as a measure of the deviation between the simulated mean and the observed statistic. Averaging this quantity across all finance/finance pairs and comparing it to finance/tech pairs, for example, will summarize any patterns in the discrepancies. Note also that this statistic could be normalized by the standard deviation calculated through simulation. We choose not to do this in the results we present, but it is easy to get accurate estimates for this normalized quantity by using the rule-of-thumb estimates for the standard deviation mentioned above.

Table 9 gives the results. The standard errors reported are calculated from the standard deviation of the sample of \(C_{BV}\) values for a particular pairing (all tech/tech stocks, for example) and dividing by the square root of the number of stocks in the set. Our first observation is that the discrepancy is noticeably larger in within-industry pairs than in out-of-industry pairs, especially in comparison to the reported standard errors.\(^{15}\) Secondly, the increase we see in this discrepancy from the daily level to the monthly level is much more significant for within-industry pairs than for out-of-industry pairs. We must note that normalizing the discrepancy by the standard error of the simulated distribution suggests that it is roughly constant between daily, weekly, and monthly levels for within-industry pairs and that it decreases with time for out-of-industry pairs. These results suggest that our simulations are likely missing some systematic effect that relates stocks in the same industry. Determining this missing factor would be an interesting extension to these results.

The observations that we make in this section are:

- With parameters estimated from the data, the simulations do manage to replicate the pattern that \(C_{BV}^{(d)} < C_{BV}^{(w)} < C_{BV}^{(m)}\).
- However, this increase is often not nearly as strong in the simulations as in the observed data.
- The simulated moments agree more closely with the observed data when considering pairs of stocks within the same industry than when considering pairs from two separate industries.

The final result, if it proves to be significant under more formal statistical tests, is perhaps the most interesting since it suggests a clear direction for future investigation.

6 Conclusion

The majority of this paper has dealt with empirical investigations of of a statistic that has not been carefully considered in the literature: the correlation between two time series of the relative contribution of jumps \(RJ_t\).

\(^{15}\)Note that a naive \(t\)-test would not suffice in this setup. BAC, for example, is used in a number of the finance/finance comparisons along with in the food/finance comparisons. The samples are not independent. However, judging from the small standard errors, we conjecture that any reasonable statistical test would probably return that the means between the finance/finance and food/finance groups are significantly different.
This study was motivated by the findings that this correlation is larger for stocks within the same industry than it is for stocks from separate industries and that this correlation has a tendency of increasing if we calculate $R_{J_1}$ at the weekly or monthly level instead of at the daily level. This increase is most noticeable for pairs of stocks from the same industry. Very basic Monte Carlo simulations indicate that much of this correlation may be an artifact of a high diffusive volatility between a pair of stocks—a somewhat surprising (albeit explicable) result given that $R_{J_1}$ is designed to measure the contribution of the jumps to the total variance. However, quantities such as the magnitude of the jumps and the frequency of cojumps seem to have a noticeable effect on this correlation as well. Furthermore, running Monte Carlo simulations where parameters are estimated directly from the data successfully replicates some of the empirical features we discovered, but the discrepancy between the simulated results and the observed data is greatest when we consider stocks in the same industry.

A first consideration is to improve the Monte Carlo simulation method. We do not include many aspects that are used in state-of-the-art simulations. For example, Bollerslev, Todorov, and Li (2011) use a much more sophisticated simulation method that incorporates microstructure noise, a stochastic volatility model, and properly correlated cojumps from a truncated normal distribution. While it is unlikely that including microstructure noise or changing the distribution of the jumps will explain away the discrepancies between the simulation and the data, stochastic volatility may be an important part of the puzzle. While an increase in volatility (perhaps due to volatility clustering together for all finance stocks, for example) in both stocks in the pair does not imply that $R_{J_1}$ will be larger for both stocks, it may have an unanticipated effect. At a more general level, it would be interesting to understand whether the discrepancy between data and simulation is endemic to even the most sophisticated models: in such a case, the patterns we observe in $C_{BV}$ may point to a characteristic in the data that has been overlooked.

Another possible explanation for the large discrepancy within-industry (compared to the smaller one out-of-industry) is that the underlying frequency of jumps may change with time, and this pattern may be related for stocks in the same industry. For example, if all finance stocks undergo a period of a few months where even idiosyncratic jumps are more frequent (if not cojumps), then we could explain the ordering $C_{BV}^{(d)} < C_{BV}^{(w)} < C_{BV}^{(m)}$. A more careful analysis of univariate $R_{J_1}$ series along with other direct estimates of the frequency of jumps could lend some insight into this hypothesis.

However, another plausible conclusion is that the $C_{BV}$ statistic is too “noisy” to use for any economically meaningful analysis. Its nonlinear nature makes analytic results intractable, so little progress can easily be made on analyzing properties of this statistic without the aid of simulation. We have also observed that many different parameter values often give rise to the same values of $C_{BV}$. While a result that even a state-of-the-art simulation cannot replicate this moment in the data (using reasonable parameter estimates)
would certainly be interesting, it may be difficult to glean any information from $C_{BV}$ that could not be determined through other, more direct, methods.

A Complications with Using a Normal Model for Jumps

As discussed in Section 4.1, we assume that the magnitude of jumps is normally distributed with mean 0 and that the magnitude of cojumps is jointly normal, also with mean 0. An issue with such an approach is that it assumes that we see many small “jumps” with magnitude close to zero. Such movements should likely not be categorized as jumps and will certainly not be distinguished from the diffusive price process using jump-detection tests. As a result, such an assumption may be unrealistic.

However, if the true process does involve jump magnitudes that are normally distributed, then we can easily determine the sign of the bias in the estimation procedure described in Section 4.3. The Jiang-Oomen test will not identify days containing small jumps as “jump days,” and our estimates for $\lambda_i$ and $\lambda_c$ will be biased downward. If we employ the procedure proposed by Andersen, Bollerslev, and Dobrev (2006), then we will identify only the large jumps as jumps. As a simple model of this phenomenon, consider sampling numbers $X_i$ from a standard normal distribution but suppose further that we only observe $X_i$ if $|X_i| > c$ for some cutoff value $c$. If we then use the sample standard deviation as an estimate for the standard deviation of the underlying data generating process—the standard normal—we will clearly obtain a value greater than 1. Thus, our estimates for $\sigma_{cj,i}$ and $\sigma_{jump,i}$ are biased upward.

Finally, we can use the same model to predict the sign of the bias in our estimate of $\rho_j$. Imagine we generate data $(X_i, Y_i)$ from a multivariate normal where the marginal distributions are standard normal and the underlying correlation is $\rho$. If we only observe data points where $|X_i| > c$ and $|Y_i| > c$ and use the sample correlation as an estimate for $\rho$, we will overestimate the magnitude of $\rho$. The estimator we calculate will be

$$
\frac{\mathbb{E}[XY|X > c \cap Y > c]}{\text{sd}(X|X > c) \cdot \text{sd}(Y|Y > c)}.
$$

(25)

Figure 8 shows the estimated correlation, given by this equation, for various values of the cutoff $c$ and the true underlying correlation $\rho$. The values in Figure 8 are calculated by numerically evaluating the integral corresponding to Equation 25. We note that the bias becomes more severe as the cutoff increases, as we would expect. Furthermore, as we would anticipate from symmetry arguments, the bias when the true correlation is $-1, 0$, or $1$ is 0.
### B Tables

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Industry</th>
<th>Start Date</th>
<th>End Date</th>
<th>Number of Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAPL</td>
<td>Tech</td>
<td>4/16/97</td>
<td>12/30/10</td>
<td>3419</td>
</tr>
<tr>
<td>BAC</td>
<td>Finance</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3422</td>
</tr>
<tr>
<td>BK</td>
<td>Finance</td>
<td>4/8/97</td>
<td>12/30/10</td>
<td>3418</td>
</tr>
<tr>
<td>C</td>
<td>Finance</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3420</td>
</tr>
<tr>
<td>CMCSA</td>
<td>Tech</td>
<td>4/15/97</td>
<td>12/30/10</td>
<td>3418</td>
</tr>
<tr>
<td>COF</td>
<td>Finance</td>
<td>4/8/97</td>
<td>12/30/10</td>
<td>3410</td>
</tr>
<tr>
<td>CPB</td>
<td>Food/Agricultural</td>
<td>4/8/97</td>
<td>12/30/10</td>
<td>3419</td>
</tr>
<tr>
<td>CSCO</td>
<td>Tech</td>
<td>4/16/97</td>
<td>12/30/10</td>
<td>3420</td>
</tr>
<tr>
<td>DELL</td>
<td>Tech</td>
<td>4/15/97</td>
<td>12/30/10</td>
<td>3418</td>
</tr>
<tr>
<td>GOOG</td>
<td>Tech</td>
<td>8/20/04</td>
<td>12/30/10</td>
<td>1591</td>
</tr>
<tr>
<td>GS</td>
<td>Finance</td>
<td>5/5/99</td>
<td>12/30/10</td>
<td>2907</td>
</tr>
<tr>
<td>HNZ</td>
<td>Food/Agricultural</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3419</td>
</tr>
<tr>
<td>HPQ</td>
<td>Tech</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3422</td>
</tr>
<tr>
<td>IBM</td>
<td>Tech</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3423</td>
</tr>
<tr>
<td>INTC</td>
<td>Tech</td>
<td>4/16/97</td>
<td>12/30/10</td>
<td>3417</td>
</tr>
<tr>
<td>JPM</td>
<td>Finance</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3422</td>
</tr>
<tr>
<td>KFT</td>
<td>Food/Agricultural</td>
<td>6/13/01</td>
<td>12/30/10</td>
<td>1950</td>
</tr>
<tr>
<td>KO</td>
<td>Food/Agricultural</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3421</td>
</tr>
<tr>
<td>MON</td>
<td>Food/Agricultural</td>
<td>10/18/00</td>
<td>12/30/10</td>
<td>2456</td>
</tr>
<tr>
<td>MSFT</td>
<td>Tech</td>
<td>4/16/97</td>
<td>1/25/10</td>
<td>3184</td>
</tr>
<tr>
<td>T</td>
<td>Tech</td>
<td>4/9/97</td>
<td>12/30/10</td>
<td>3422</td>
</tr>
<tr>
<td>XOM</td>
<td>Energy</td>
<td>12/1/99</td>
<td>12/30/10</td>
<td>2763</td>
</tr>
</tbody>
</table>

Table 1: Stocks used in the analysis, along with start and end dates and the number of days included in the data.
## Table 2: Mean and standard errors for $R_{20}$, calculated by year and by stock.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>APL</td>
<td>0.0763</td>
<td>0.137</td>
<td>0.183</td>
<td>0.211</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
<td>0.0763</td>
<td>0.137</td>
<td>0.183</td>
<td>0.211</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
</tr>
<tr>
<td>BAC</td>
<td>0.0966</td>
<td>0.043</td>
<td>0.043</td>
<td>0.159</td>
<td>0.159</td>
<td>0.263</td>
<td>0.127</td>
<td>0.127</td>
<td>0.043</td>
<td>0.043</td>
<td>0.159</td>
<td>0.159</td>
<td>0.263</td>
<td>0.127</td>
</tr>
<tr>
<td>BP</td>
<td>0.0673</td>
<td>0.067</td>
<td>0.067</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td>0.067</td>
<td>0.067</td>
<td>0.067</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
</tr>
<tr>
<td>C</td>
<td>0.0068</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>CMCSA</td>
<td>0.159</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
<td>0.0763</td>
<td>0.137</td>
<td>0.183</td>
<td>0.211</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
</tr>
<tr>
<td>COF</td>
<td>0.0063</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>CPB</td>
<td>0.0063</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>CS</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>DELL</td>
<td>0.0063</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>GOOG</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>GS</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>HNZ</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>HPQ</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>IBM</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>INTC</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>JPM</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>KO</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>T</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>XOM</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
<td>0.053</td>
<td>0.115</td>
<td>0.172</td>
<td>0.201</td>
<td>0.172</td>
<td>0.172</td>
<td>0.115</td>
</tr>
</tbody>
</table>

Overall:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>0.0763</td>
<td>0.137</td>
<td>0.183</td>
<td>0.211</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
<td>0.0763</td>
<td>0.137</td>
<td>0.183</td>
<td>0.211</td>
<td>0.211</td>
<td>0.183</td>
<td>0.137</td>
</tr>
<tr>
<td>Food</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
</tr>
<tr>
<td>Tech</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
</tr>
<tr>
<td>All</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
<td>0.109</td>
</tr>
</tbody>
</table>

Note: The table above presents the mean and standard errors for $R_{20}$, calculated for each stock and year. The data is presented for 14 stocks, with the average for the overall market also shown. The results are calculated using 5-minute increments.
Frequency parameter of idiosyncratic jumps for stock $i$. Constant across time.

$\rho_d$ Diffusive correlation between two stocks.

$\rho_c$ Correlation in cojump size

$\sigma_{jump,i}$ Standard deviation of idiosyncratic jump size for stock $i$

$\sigma_{cj,i}$ Standard deviation of cojump size for stock $i$

$\lambda_i$ Frequency parameter of idiosyncratic jumps for stock $i$ in day$^{-1}$

$\lambda_c$ Frequency parameter of idiosyncratic jumps for stock $i$ in day$^{-1}$

Table 3: Average values for $C_{BV}$ separated into industries along with standard errors. For example, the first entry in the table is the average $C_{BV}$ of AAPL and all other tech stocks. The table is separated into industries as well. The general observation is that two stocks in the same industry tend to have a higher value of $C_{BV}$ than a pair of stocks from different industries. This observation is most apparent with food and finance stocks.

Table 4: Notation summary and estimation methods
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_i$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\lambda_i$</td>
<td>0.25</td>
</tr>
<tr>
<td>$\sigma_c \equiv \sigma_{cj,i} = \sigma_{jump,i}$</td>
<td>0.025</td>
</tr>
<tr>
<td>$\rho_d$</td>
<td>0, 0.20, 0.40, 0.60, 0.80</td>
</tr>
<tr>
<td>$\rho_c$</td>
<td>0, 0.25, 0.50, 0.75</td>
</tr>
<tr>
<td>$\lambda_c$</td>
<td>0.075, 0.25, 1</td>
</tr>
</tbody>
</table>

Table 5: Parameter values used for simulations in Section 4.2. Units are given in Table 4.

<table>
<thead>
<tr>
<th>Pairs</th>
<th>$\rho_d$</th>
<th>$\rho_c$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\sigma_{jump,1}$</th>
<th>$\sigma_{jump,2}$</th>
<th>$\sigma_{cj,1}$</th>
<th>$\sigma_{cj,2}$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\lambda_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC/GS</td>
<td>0.396</td>
<td>0.928</td>
<td>0.0235</td>
<td>0.0216</td>
<td>0.0184</td>
<td>0.0174</td>
<td>0.0162</td>
<td>0.0188</td>
<td>0.0450</td>
<td>0.0616</td>
<td>0.0065</td>
</tr>
<tr>
<td>DELL/IBM</td>
<td>0.383</td>
<td>0.949</td>
<td>0.0236</td>
<td>0.0157</td>
<td>0.0144</td>
<td>0.0115</td>
<td>0.0383</td>
<td>0.0278</td>
<td>0.0509</td>
<td>0.0514</td>
<td>0.0056</td>
</tr>
</tbody>
</table>

Table 6: Estimated parameters for two pairs of stocks. The subscript 1 refers to the first stock and 2 to the second. The values $\sigma_1$ and $\sigma_2$ are given at the daily level; the other volatilities are per jump. The frequency parameters $\lambda$ are in units day$^{-1}$.

<table>
<thead>
<tr>
<th>$C_{BV}^{\text{sim}}$</th>
<th>Daily</th>
<th>Weekly</th>
<th>Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0553</td>
<td>0.0815</td>
<td>0.0857</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.091)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>$C_{BV}^{\text{data}}$</td>
<td>0.0658</td>
<td>0.18</td>
<td>0.305</td>
</tr>
<tr>
<td>$C_{\text{MinV}}^{\text{sim}}$</td>
<td>0.0451</td>
<td>0.07</td>
<td>0.0775</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.078)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>$C_{\text{MinV}}^{\text{data}}$</td>
<td>0.0793</td>
<td>0.153</td>
<td>0.222</td>
</tr>
<tr>
<td>$C_{\text{MedV}}^{\text{sim}}$</td>
<td>0.0553</td>
<td>0.0799</td>
<td>0.0869</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.088)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>$C_{\text{MedV}}^{\text{data}}$</td>
<td>0.0825</td>
<td>0.173</td>
<td>0.249</td>
</tr>
</tbody>
</table>

Table 7: Comparison of simulated and observed values of $C_*$ for BAC and C at the daily, weekly, and monthly levels. The standard error given in the simulation row is the standard deviation of the distribution of $C_*$ calculated in the simulation, using the 1000 sample paths. There is a noticeable systematic discrepancy between the simulations and the data in that the simulation moments tend to be lower than the observed ones—especially at weekly and monthly levels.

<table>
<thead>
<tr>
<th>$C_{BV}^{\text{sim}}$</th>
<th>Daily</th>
<th>Weekly</th>
<th>Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0554</td>
<td>0.105</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.087)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>$C_{BV}^{\text{data}}$</td>
<td>0.0605</td>
<td>0.137</td>
<td>0.144</td>
</tr>
<tr>
<td>$C_{\text{MinV}}^{\text{sim}}$</td>
<td>0.0373</td>
<td>0.0876</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.08)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>$C_{\text{MinV}}^{\text{data}}$</td>
<td>0.0469</td>
<td>0.112</td>
<td>0.0582</td>
</tr>
<tr>
<td>$C_{\text{MedV}}^{\text{sim}}$</td>
<td>0.0532</td>
<td>0.105</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.085)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>$C_{\text{MedV}}^{\text{data}}$</td>
<td>0.0663</td>
<td>0.12</td>
<td>0.0972</td>
</tr>
</tbody>
</table>

Table 8: Comparison of simulated and observed values of $C_*$ for BK and KFT at the daily, weekly, and monthly levels. The standard error given in the simulation row is the standard deviation of the distribution of $C_*$ calculated in the simulation, using the 1000 sample paths. Unlike in Table 7, there is no systematic discrepancy between the simulations and the data.
Table 9: Average values for \( |C_{BV}^{\text{sim}} - C_{BV}^{\text{data}}| \) at daily, weekly, and monthly levels to quantify the discrepancy between simulations and the moments in the observed data. This discrepancy is averaged over all within-industry pairs as well as all out-of-industry pairs. Other statistics, such as \( C_{MinV} \) and \( C_{MedV} \), as well as those calculated using \( \rho \) and \( \tau \), show similar patterns.
C Figures

Figure 1: Sample plots of $R_J$ using KO. (a) The daily $R_{JBV}$ series plotted against time. The series is calculated using 5-minute intervals. (b) Kernel density plots of $R_J$ series using various jump-robust statistics.

Figure 2: Values of $C_{BDV}^{(d)}$, $C_{BDV}^{(w)}$, and $C_{BDV}^{(m)}$ for BAC and DELL. The points in red represent tech stocks, those in blue are finance stocks, and those in green are food/agricultural stocks. The general observations we find—more clear in the figure for BAC—is that pairs of stocks in the same industry tend to have higher values of $C_{BDV}$. Furthermore, the magnitude of $C_{BDV}$ increases with the time frame considered.
Figure 3: Comparisons of $C_{BV}^{(w)}$ vs. $C_{BV}^{(d)}$ and $C_{BV}^{(m)}$ vs. $C_{BV}^{(d)}$. The solid red line is a $45^\circ$ line. We can reject the hypothesis that the slope of the regression line through the points in either case coincides with the slope of the $45^\circ$ line.
Figure 4: Comparisons of \( JC \) at 1\% level against variations of \( C^{(d)}_{M\text{VinV}} \). In (a) and (b), the industry-specific colors (red for tech stocks, blue for finance stocks, and green for food/agricultural stocks) is used to compare these statistics against each other. Comparisons of the ticker symbols suggests that the ordering and magnitudes differ noticeably between \( JC \) and \( C^{(d)}_{M\text{VinV}} \). Panel (c) plots \( JC \) against \( \tau^{(d)}_{BV} \) for all pairs of stocks considered. We see no clear positive association between these statistics.
Figure 5: $C_{BV}$ as a function of $\rho_d$ for various values of $\sigma_c$ and $\lambda_c$, fixing $\rho_c = 0.5$. The different panels give different values of $\lambda_c$ whereas the different colors within a panel correspond to different $\sigma_c$. Standard errors on the mean values calculated through simulation are small and thus ignored in the plots. The general observation is that $\rho_d$ has a strong effect on $C_{BV}$. 
Figure 6: $C_{BV}$ as a function of $\rho_c$ for various values of $\sigma_c$ and $\lambda_c$, fixing $\rho_d = 0.4$. The different panels give different values of $\lambda_c$ whereas the different colors within a panel correspond to different $\sigma_c$. Standard errors on the mean values calculated through simulation are small and thus ignored in the plots. The general observation is that $\rho_c$ has a very small effect on $C_{BV}$, although for sufficiently small $\sigma_c$, the effect is noticeable at high $\rho_c$. 

(a) $\lambda_c = 0.075$

(b) $\lambda_c = 0.25$

(c) $\lambda_c = 1$
Figure 7: *Simulated* values of $C_{B^V}^{(d)}$, $C_{B^V}^{(w)}$, and $C_{B^V}^{(m)}$ for BAC and DELL. The color coding is the same as before: red for tech stocks, blue for finance stocks, and green for food/agricultural stocks. See Figure 2 to compare these simulated moments to the moments observed in the data.

Figure 8: Measured correlation of two jointly normal variables with correlation $\rho$ (plotted on the “Correlation” axis) when observing only the realizations that are above some cutoff $c$ (plotted on the “Cutoff” axis). The value of the function is calculated through numerical integration of Equation 25.
References


