Cross-Stock Comparisons of the Relative Contribution of Jumps to Total Price Variance

Vivek Bhattacharya

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

2012

Acknowledgements

Abstract

This paper uses high-frequency price data to study the relative contribution of jumps to the total volatility of an equity. In particular, it systematically compares the relative contribution of jumps across a panel of stocks from three different industries by computing the cross-correlation of this statistic for pairs of stocks. We identify a number of empirical regularities in this cross-correlation and compare these observations to predictions from a standard jump-diffusion model for the joint price process of two stocks. A main finding of this paper is that this jump-diffusion model, when calibrated to particular pairs of stocks in the data, cannot replicate some of the empirical patterns observed. The model predictions differ from the empirical observations systematically: predictions for pairs of stocks from the same industry are on the whole much less accurate than predictions for pairs of stocks from different industries. Some possible explanations for this discrepancy are discussed.

1   Introduction

The availability of high-frequency financial market data has spurred much research related to the volatility of the prices of stocks and other financial securities. Estimators with appealing theoretical properties have been developed that take advantage of the minute-by-minute or even tick-by-tick price data that can easily be obtained. The essential idea behind these estimators is that high-frequency price movements provide a more accurate picture of the underlying price process, and these small movements can be aggregated to generate estimates of the volatility of the stock over a particular period of time. Of course, understanding and estimating volatility is of much practical import as well. As the most common measure of risk, accurate estimates of volatility are important for options pricing, asset allocation, and risk management. Well-established models in finance, such as the Black-Scholes options pricing model (Black & Scholes, 1973) and the Markowitz mean-variance framework (Markowitz, 1952) all require estimates of the volatility of the securities as inputs to the model.

A second benefit of high-frequency data is that it allows researchers to conduct statistical inference separately on the various components of the price process. Modern models of stock prices separate the price process into two components: a smooth "diffusive" process and a discontinuous process that allows for "jumps" to occur periodically in the time series of prices.[1] The most common explanation for the presence of jumps, dating back to Merton (1976), is that they are the manifestations of an efficient market incorporating new, unanticipated information into the price essentially as soon as the new information is released. There is substantial econometric evidence for the presence of such a discontinuous component in observed data. Andersen, Benzoni, and Lund (2002) study various models of equity return data and conclude that discrete jumps must be present in any reasonable model to describe the observed data. Drost, Nijman, and Werker (1998) perform similar statistical tests on the time series of dollar exchange rates and reach a similar conclusion in this different asset class: a pure diffusive model cannot explain the data entirely. Furthermore, these jumps complicate many fundamental problems in finance: for example, analytic analysis of

---

[1]The discussion in the first part of this paragraph is based on Rognlie (2010).

4

options pricing models is difficult in the presence of jumps in the price process, and researchers have had to impose various assumptions on the jump process to make reasonable progress (Kuo, 2002). Researchers have developed many statistical techniques to test for the presence of such jumps. Some of these techniques exploit properties of the higher-order moments of the price process, which will have different behavior depending on whether a jump is present (Ait-Sahalia & Jacod, 2009; Jiang & Oomen, 2008). Another statistical test utilizes different estimators for volatility to test for the presence of jumps (Barndorff-Nielsen & Shephard, 2005).

Although this paper will briefly employ jump tests, the emphasis will be on studying the contribution of jumps to the volatility. Indeed, this analysis can be performed without directly testing for jumps at all. While realized variance, developed by Barndorff-Nielsen and Shephard (2004), estimates the quadratic variation of the total price process within a particular time period, researchers have proposed a number of simple estimators of variance that are jump-robust. These estimators are consistent for the volatility of the diffusive component of the price process. For example, Barndorff-Nielsen and Shephard (2004) propose bipower variation, and Andersen, Dobrev, and Schaumburg (2010) propose alternate other estimators, called *MinVar* and *MedVar*. All these estimators exploit the observation that large intraday price movements are likely to be due to jumps rather than to the diffusive process. Since the total quadratic variation within a day can come from either the diffusive component or the jump component, the difference between realized variance and one of the jump-robust estimators of the variance returns an estimator for the daily variance of the jump process. Dividing this quantity by realized variance estimates the *relative contribution of jumps to total price variance*, a statistic that was first considered by Huang and Tauchen (2005).

Many variations of the contribution of jumps to total price variance have been studied in the literature before for a variety of applications. The interest in the difference between realized and bipower variation began with Barndorff-Nielsen and Shephard (2004), who introduce this statistic as an estimator for the quadratic variation of the jump process and study basic properties. For ease of exposition, denote this statistic by J_t. Among other things, they conduct Monte Carlo tests and show that truncating J_t at 0 to ensure that it is positive improves finite sample behavior. Andersen,

Bollerslev, and Diebold (2007) study J_t as a component in a volatility forcasting model. They extend the HAR-RV model of Corsi (2009)—which parameterizes future realized variance as a linear combination of one-day, one-week, and one-month lagged variances—to include J_t as a regressor. They find that predicatibility in future realized volatilities comes almost entirely from the volatility of the diffusive component. Andersen et al. (2007) also conduct some basic time-series analysis of J_t for the S&P500 Futures Index[2] and find that serial autocorrelations up to the tenth order are significant. A more recent paper by Bollerslev, Kretschmer, Pigorsch, and Tauchen (2009) studies properties of the difference in the logs of realized and bipower variance, which we will denote LJ_t for ease of exposition. They find that the distribution of LJ_t calculated from the S&P Futures Index is positively skewed. Furthermore, they only find first-order and fifth-order autocorrelations in the LJ_t time series significant; an explanation for the significance of the fifth-order autocorrelation is that certain macroeconomic announcements are weekly events. Using these empirical regularities of the LJ_t series, the authors formulate a simultaneous equation model for the jump statistic LJ_t and the bipower variation and employ it to study the leverage effect.

There are certain benefits to studying the normalized difference in the realized and bipower variances by the realized variance. The interest in this normalized statistic began with its application to statistical tests for jumps. Barndorff-Nielsen and Shephard (2005) study this quantity as part of a test statistic for the presence of jumps. Huang and Tauchen (2005) use Monte Carlo simulations to verify that this test statistic has superior finite sample properties over a broad range of model specifications. Furthermore, this statistic is also easily interpretable: a number close to 1 implies that nearly all of the daily quadratic variation is due to the jump process while a number close to 0 implies that almost all of the daily quadratic variation is due to the diffusive process. In contrast, it is more difficult to put an increase in the unnormalized estimator J_t into context without also studying the increase in the realized variation. This increase in interpretability comes at a technical cost: by introducing a ratio into the statistic, many mathematical manipulations become difficult.

This paper studies an aspect of the relative contribution of jumps that has not been considered

---

[2]For the purposes of this discussion, this index can be viewed as a summary of the entire market.

before in the literature: it compares the time series of the relative contribution of jumps across a panel of many stocks using high-frequency data. In particular, it computes the cross-correlation of the relative contribution of jumps of the time-series of pairs of stocks and notes a number of empirical regularities. This paper focuses on comparing properties of the relative contribution of jumps for pairs from the same industry to pairs from different industries. Such exploratory data analysis can add to the current understanding of high-frequency equity data. For example, these empirical findings may be interesting and relevant to researchers studying the relative contribution of jumps in a portfolio setting with stocks from various industries. This work then considers the cross-correlation statistic in the context of a simple but standard model of the price process of a pair of stocks and asks whether the model, properly calibrated, can replicate the observed values. Simple statistical methods are used to compare the model predictions to the empirical observations.

The paper is organized as follows. Section 2 provides the theoretical background, describing the price process as well as the estimators for variance. Section 3 discusses the details of jump tests and also formally introduces the cross-correlation of the relative contribution of jumps. It then presents the model used in this paper and describes the statistical methodology employed to calibrate the model to the data. It is difficult to make analytic progress even with this simple model, so Section 4 conducts Monte Carlo experiments to test the effects of the underlying parameters of the model on the statistic. The next three sections discuss the empirical results of the study. The data, along with empirical considerations such as microstructure noise, are described in Section 5. Section 6 documents the empirical observations related to the cross-correlation statistic and concludes by discussing the relation between these empirical observations and the theoretical findings from the previous section. Section 7 calibrates the model directly to particular pairs of stocks in the data and compares the model predictions with the empirical observations. Sections 8 and 9 conclude.

2 Theoretical Background

2.1 Continuous Time Models for Returns

The standard model of a stock price process is that the efficient price $p(t)$ follows the stochastic differential equation

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

where $\mu(t)$ and $\sigma(t)$ are functions of time that 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, $\mu(t)$ is independent of $w(t)$, although it is common in many modern models to assume that $\sigma(t)$ is dependent on the Wiener increment. The dollar price of the stock is $P(t) = \exp{(p(t))}$, and as a result, Equation (1) is an assumption that the efficient price of a security follows a process called geometric Brownian motion. Furthermore, this process has the property that all sample paths, or particular realizations of this process, are almost surely continuous. This geometric Brownian motion model has many benefits, perhaps the most important of which is its analytical tractability. For example, the Black-Scholes equation for options pricing obtains an explicit formula for the value of an option if the underlying price process of the security can be described by Equation (1). As a result, initial studies of the behavior of stock prices often used Equation (1) or a simple variant as the basis of their mathematical framework.

The modern literature on the stock market, however, has discredited the assumption that the observed sample paths are continuous, instead suggesting that discontinuites play an essential role in the process. As discussed in the introduction, these discontinuities, or "jumps," have been observed in many asset classes and have a significant effect on many aspects of finance, such as securities pricing. Mathematically, an additional term $dJ(t)$ allows for jumps in the price process. The differential equation used in this paper, therefore, is

$$dp(t) = \mu(t)dt + \sigma(t)dw(t) + dJ(t), \tag{2}$$

8

where $J(t) - J(s) = \sum_{s \leq t' \leq t} \kappa(t')$ and $\kappa(\cdot)$ is the magnitude of a jump. It is possible to write $dJ(t) = \kappa(t)dq(t)$, where $dq(t)$ is a binary variable that is 1 if there is a jump at time t and equal to 0 otherwise. A goal of financial econometrics is to use observed prices to conduct inference on the jump component separately from the continuous component.

2.2 Estimators for the Variance

Estimating properties of the volatility of a security is an important aspect of econometric analysis of financial times series. Equation (2) indicates that variance in the more general price process can come from two components: the diffusive volatility process $\sigma(t)$ and the jumps $\kappa(t)$. A parameter that aggregates the contibutions from both these processes is the quadratic variation on day t, defined as

$$QV_t \equiv \int_{t-1}^{t} \sigma(s)^2 \, ds + \sum_{t-1 \leq s \leq t} \kappa(s)^2. \tag{3}$$

A common nonparametric estimator for QV_t is the realized variance

$$RV_t \equiv \sum_{j=1}^{M} r_{t,j}^2, \tag{4}$$

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 $(j-1)^{\text{st}}$ and j^{th} price observations in a day, and M is the number of returns observed in a single day. Barndorff-Nielsen and Shephard (2004) show that RV_t is consistent[3] for the quadratic variation QV_t as M increases. Of course, a larger value of M implies that returns are observed more frequently during the day. While deriving the asymptotic properties of RV_t is mathematically difficult, the intuition is straightforward: the square of the geometric return is an estimate of the variance in a particular interval, and adding the variances calculated from a sequence of small intervals gives an estimate of the daily variance. Since the geometric return does not discriminate

---

[3]A sequence of random variables $\{X_n\}$ is said to converge in probability to a (possibly random) value X if for every $\epsilon > 0$, $\Pr(|X_n - X| > \epsilon) \to 0$, where $\to$ represents convergence as a limit of real numbers. An estimator is (asymptotically) consistent for a population parameter if it converges in probability to the parameter as more data is collected. Essentially, consistency means that an estimator is statistically reasonable is the sense that it gets closer to the true value if using more data.

between whether the price change comes from the continuous component or the jump component, RV_t is an estimator for the total variance of the process.

It is often useful to study properties of $\sigma(t)$ individually and isolate it from the jump component. The appropriate population parameter in this case is the first term in Equation (3), which is called the integrated variance and denoted $IV_t \equiv \int_{t-1}^{t} \sigma(s)^2 \, ds$. It must be stressed that IV_t is the variation coming entirely from the continuous component of the price. To estimate IV_t, Barndorff-Nielsen and Shephard propose the bipower variation estimator

$$BV_t \equiv \mu_1^{-2} \cdot \frac{M}{M-1} \cdot \sum_{j=2}^{M} |r_{t,j-1}| \, |r_{t,j}| , \tag{5}$$

where $\mu_k \equiv \mathbb{E}\left(|Z|^k\right)$ if Z is standard normal. Andersen et al. (2010) propose the alternate estimators

$$MinVar_t \equiv \frac{\pi}{\pi-2} \cdot \frac{M}{M-1} \cdot \sum_{j=2}^{M} \min\left\{|r_{t,j-1}|, |r_{t,j}|\right\}^2 \quad \text{and} \tag{6}$$

$$MedVar_t \equiv \frac{\pi}{6 - 4\sqrt{3} + \pi} \cdot \frac{M}{M-2} \cdot \sum_{j=3}^{M} \mathrm{med}\left\{|r_{t,j-2}|, |r_{t,j-1}|, |r_{t,j}|\right\}^2 . \tag{7}$$

The intuition behind each of these estimators can be made clear. A price change $r_{t,j}$ is likely to be a jump instead of a continuous movement if it is considerably larger in magnitude than other nearby movements in the same day. Each of these three estimators tries to limit the contribution of such large price changes to the variance: BV dilutes the effect of large changes by multiplying them with what is likely a smaller value while $MinVar$ and $MedVar$ effectively replace these large values with smaller neighboring values. It can be shown that all three estimators converge in probability to IV_t and thus allow researchers a way to separate the volatility from the diffusive component from the volatility from the jump component.

Given these estimators, it is easy to devise a consistent estimator for the ratio $(QV_t - IV_t)/QV_t$ of the total volatility due to the jumps. Letting $\times$ be any of the three jump robust statistics described

above, an estimator for this ratio is

$$RJ_{\times,t} \equiv \frac{RV_t - \times_t}{RV_t}. \tag{8}$$

While $IV_t \leq QV_t$, it is not necessarily the case that $\times_t \leq RV_t$ in finite samples;[4] thus, $RJ_{\times,t}$ may be negative in finite samples. Andersen et al. (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 values of $RJ_{\times,t}$ if necessary. The rationale is that empirical investigations suggest that the statistics this paper considers are not altered much by the truncation.

3 Statistical Methods

3.1 Jump Tests

Based on the statistics described above, 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. A common test in the literature, and one that exemplifies the usefulness of $RJ_{BV,t}$ in statistical inference, is proposed by Barndorff-Nielsen and Shephard (2005). Huang and Tauchen (2005) modify this test statistic slightly and propose

$$Z = \frac{RJ_{BV,t}}{\sqrt{\left(\frac{\pi^2}{4} + \pi + 5\right) \cdot \frac{1}{n} \cdot \max\left\{1, \frac{\widehat{Q}_t}{BV_t^2}\right\}}}, \tag{9}$$

where $\widehat{Q}$ is a consistent estimator for the integrated quarticity $\int_{t-1}^{t} [\sigma(s)]^4 ds$, and $\mu_{4/k}$ is as defined in the previous section. The two estimators Tauchen and Huang consider are the tripower quarticity

---

[4]It is easy to devise a sequence of real numbers $\{r_i\}_{i=1}^{n}$ such that $\sum_{i=1}^{n} r_i^2 < \mu_1^{-2} \cdot [n/(n-1)] \cdot \sum_{i=2}^{n} |r_{i-1}r_i|$; simply consider the case where $r_i = 1$ for all i. Note that there exists an $\epsilon > 0$ such that the inequality holds whenever $r_i \in (1 - \epsilon, 1 + \epsilon)$ for all i. Even under the assumption that r_i is drawn from a normal distribution, such an outcome happens with positive probability. Thus, we have found an example that shows that in finite samples, it is not necessarily the case that $BV_t < RV_t$. Similar examples can be found for both $MinVar$ and $MedVar$.

and the quadpower quarticity, given by

$$\widehat{Q}_k = M \cdot \mu_{4/k}^{-k} \cdot \frac{M}{M-k+1} \sum_{i=k+1}^{M} \prod_{j=1}^{k} |r_{i-j}|^{4/k},\tag{10}$$

for $k = 3$ and $k = 4$, respectively. The test statistic is asymptotically standard normal under the null hypothesis, and the null hypothesis of no jumps is thus rejected if $Z > \Phi^{-1}(1-\alpha)$, where α is the significance level of the test and Φ is the cumulative distribution function of the standard normal. Huang and Tauchen's Monte Carlo simulations support the use of tripower quarticity in finite samples. The rationale behind the B-NS test is that a large difference in RV_t and BV_t (and thus a large value of $RJ_{BV,t}$) will be due to jumps in the price process. Therefore, this statistic provides a natural way of testing for the presence of jumps. This paper, however, aims to explore the magnitude of $RJ_{BV,t}$ and compare that to the results of jump tests. As a result, using the B-NS test will be problematic, since any artifact that affects $RJ_{BV,t}$ in finite samples will also affect the outcome of the jump test.

This paper considers a test from Jiang and Oomen (2008) that exploits higher moments of the return process, thereby providing a method of testing for jumps without resorting to $RJ_{BV,t}$. The rationale behind using this test instead of the B-NS test is that this paper is interested in $RJ_{BV,t}$ as a statistic by itself, and it is necessary to have a method of detecting jumps that does not depend on $RJ_{BV,t}$. Jiang and Oomen note that if a process is continuous, the higher moments should be small whereas in the presence of jumps, these moments will be nontrivial. The statistic Jiang and Oomen consider is the swap variance and is calculated as

$$SwV_t \equiv 2 \cdot \sum_{i=2}^{M} (R_{t,i} - r_{t,i}),\tag{11}$$

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

$$R_{t,i} \equiv \frac{P_{t,i} - P_{t,i-1}}{P_{t,i-1}} = \frac{P_{t,i}}{P_i} - 1 = \exp\left(p_{t,i} - p_{t,i-1}\right) - 1 = \exp\left(r_{t,i}\right) - 1.$$

A Taylor expansion of the exponential function shows that $SwV - RV$ contains only terms of order $r_{t,i}^2$ and higher. Since the geometric return $r_{t,i}$ over a small interval is already small in the absence of jumps, the statistics SwV and RV are approximately equal under the null hypothesis that there are no jumps. One method of studentizing the statistic $SwV - RV$ yields

$$JO \equiv M \cdot \frac{BV_t}{\sqrt{\widehat{\Omega}_{SwV}}} \left(1 - \frac{RV_t}{SwV_t}\right), \tag{12}$$

where $\widehat{\Omega}$ is a consistent estimator for the integrated sexticity $\int_{t-1}^{t}[\sigma(s)]^6 ds$. Possible estimators for this quantity include

$$\widehat{\Omega}_{SwV}^{(p)} = M^2 \cdot \frac{\mu_6}{9} \cdot \mu_{6/p}^{-p} \cdot \frac{M^2}{M - p + 1} \sum_{j=p+1}^{M} \prod_{k=1}^{p} |r_{t,j-k}|^{6/p}, \tag{13}$$

with $p = 4$ and 6 being the obvious choices. The statistic JO is asymptotically standard normal and thus lends itself to a statistical test for the presence of jumps, where the critical region is calculated through standard methods.

3.2 Cross-Correlation Structure of the Relative Contribution of Jumps

This work primarily studies basic time series properties of $RJ_{\times,t}$ and focuses specifically on analyzing pairs of these time series. While the statistics used are not complicated, introducing certain notation will improve the clarity of the paper. For a particular jump-robust estimator $\times$ and two stocks labeled 1 and 2, let the time series of $RJ_{\times,t}$ be denoted $RJ_{\times,t}^{(1)}$ and $RJ_{\times,t}^{(2)}$, respectively. The cross-correlation of these two time series will be denoted

$$C_\times \equiv \text{corr}\left(RJ_{\times,t}^{(1)}, RJ_{\times,t}^{(2)}\right). \tag{14}$$

The Pearson correlation used in Equation (14) only measures the strength of the linear relation between the relative contribution of jumps of the two stocks. It is also possible to use nonparametric

estimators of correlation, such as Spearman's ρ or Kendall's τ. These estimators measure how well the relation between the relative contribution of jumps for two stocks is explained by any monotonic function. The notation $\rho_\times$ and $\tau_\times$ will be used in this paper when considering these nonparametric estimators for correlation. However, this paper will not discuss these estimators often, as a general result is that using these statistics does not alter the results considerably. The bulk of this paper uses the standard Pearson expression for correlation when calculating cross-correlation in Equation (14).

On the other hand, the interval of time used to calculate $RJ_{\times,t}$ strongly affects the calculated cross-correlation. The *cross-correlation of the daily relative contribution of jumps* is determined by first calculating the time series $RJ_{\times,t}$ on a day-by-day basis and then calculating the cross-correlation. This statistic will be denoted $C_\times^{(d)}$. Similarly, the *cross-correlation of the weekly relative contribution of jumps* uses $RJ_{\times,t}$ calculated on a weekly basis by adding RV_t and $\times_t$ for the five days to determine the realized variation and the jump-robust measure of variation for the entire week. This statistic will be denoted $C_\times^{(w)}$. Similarly, it is possible to consider the statistic $C_\times^{(m)}$ calculated at the monthly level. The correlation calculated over different time intervals will be a key part of this paper.

It is also possible to consider a simpler version of the aforementioned statistics that only considers the presence of jumps. For a stock labeled 1, let $\mathrm{Jump}_t^{(1)}$ be the event that a jump is detected in the stock's price process in day t. The Jiang-Oomen test at either a 1 percent or a 0.1 percent significance level provides a method of estimating the indicator $\mathbb{1}\left[\mathrm{Jump}_t^{(i)}\right]$. The correlation

$$JC \equiv \mathrm{corr}\left(\mathbb{1}\left[\mathrm{Jump}_t^{(1)}\right], \mathbb{1}\left[\mathrm{Jump}_t^{(2)}\right]\right). \tag{15}$$

in the indicators suggests another method of studying the relation between the jump components of two stocks. The statistic JC abstracts away from any considerations about the magnitudes of the jumps; it does not discriminate between large jumps and small jumps and only considers whether a jump is detected by the statistical test. Note that the Pearson coefficient is numerically identical to all nonparametric measures of correlation when considering sets of indicator functions. Comparing JC to $C_\times$ is another important aspect of the empirical analysis conducted in this paper.

3.3 Model Simulations

Since analytical approaches to analyzing the statistic $C_\times$ are likely intractable,[5] this paper turns to Monte Carlo simulations of a simple model to gain more insight into the statistic. In particular, this paper considers a model of a stock price with jumps given by

$$dp^{(i)}(t) = \sigma^{(i)}dw^{(i)}(t) + dIJ^{(i)}(t) + dCJ^{(i)}(t), \tag{16}$$

where i indexes a particular stock. Like Equation (2), this model contains both a diffusive component and a jump component, but there are a few important simplifications. First, the volatility process $\sigma(t)$ of the diffusive component is taken to be independent of time. Second, the drift $\mu(t)$ is assumed to be 0. Finally, the jump process $dJ(t)$ is separated into two components: an idiosyncratic component $dIJ^{(i)}(t)$ and a common jump component $dCJ^{(i)}(t)$. Of course, the idea of this model is that price processes will be generated from the marginal distribution given in Equation (16), but relations in the joint distributions of $(dw^{(1)}, dw^{(2)})$ and $(dCJ^{(1)}, dCJ^{(2)})$ will allow for dependencies between the prices of two separate stocks.

The distributions for both the Wiener increments and the jump components are derived from simple normal distributions. The assumption concerning the diffusive components is that the underlying Brownian motions are correlated with correlation ρ_d. That is, say $\left(Z_t^{(1)}, Z_t^{(2)}\right)'$ is independent and identically distributed (iid) with joint distribution

$$\begin{pmatrix} Z_t^{(1)} \\ Z_t^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_d \\ \rho_d & 1 \end{pmatrix}\right). \tag{17}$$

Next, the three jump processes are mutually independent. The idiosyncratic jump process is defined as $dIJ_t^{(i)} = \kappa_t^{(i)}dq_t^{(i)}$ for $i \in \{1, 2\}$. The jump magnitude $\kappa^{(i)}(t)$ is distributed $N\left(0, \left(\sigma_{\text{jump}}^{(i)}\right)^2\right)$, and $q_t^{(i)}$ is a Poisson counting process with parameter $\lambda^{(i)}$. This common jump component dCJ_t is a vector $\left(dCJ_t^{(1)}, dCJ_t^{(2)}\right)' = \kappa_c(t)dq_c(t)$ where $q_c(t)$ is a Poisson counting process with parameter

---

[5]See Appendix B for a preliminary attempt.

λ_c and

$$\kappa_c(t) \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \left(\sigma_{cj}^{(1)}\right)^2 & \rho_c\sigma_{cj}^{(1)}\sigma_{cj}^{(2)} \\ \rho_c\sigma_{cj}^{(1)}\sigma_{cj}^{(2)} & \left(\sigma_{cj}^{(2)}\right)^2 \end{array}\right)\right). \tag{18}$$

Note that $dq_c(t)$ is a scalar-valued process, meaning that dCJ_t indeed induces jumps in the two price processes at the same time. If ρ_c is set to 0, these common jumps would have uncorrelated magnitudes; for positive ρ_c, these jumps not only occur at the same time but are more likely to have comparable magnitudes. Of course, all the jump magnitudes are mutually independent, and all counting processes are also mutually independent.

Using the distributions given in the previous paragraph, it is easy to simulate the price process using the Euler scheme. That is, the change in the price during a particular interval is

$$\Delta p_t^{(i)} = \sigma^{(i)} Z_t^{(i)} + dIJ_t^{(i)} + dCJ_t^{(i)}, \tag{19}$$

where $\sigma^{(i)}$ is the volatility of the diffusive component of stock i on that interval.[6] For a particular set of parameters, 1,000 sample paths are generated each with 252×5 trading days. Furthermore, prices are recorded at 5-minute increments to emulate the treatment of the data. It is clear from the above discussion that this model has many free parameters, and Table 4 summarizes the notation of this section.

Monte Carlo simulations can help understand the effect of these free parameters on the output of the model, but the large parameter space must be simplified in order to make this analysis computationally tractable. To accomplish this simplification, the parameter studies in this paper consider cases where $\sigma^{(1)} = \sigma^{(2)}$, $\lambda^{(1)} = \lambda^{(2)}$, and $\sigma_{\text{jump}}^{(i)} = \sigma_{cj}^{(i)} \equiv \sigma_c$. The parameters used for the simulations are tabulated in Table 5.

---

[6]A scaling factor is important here. Suppose there are M intervals in a given day. Set $\Delta \equiv 1/M$. If the daily volatility of a stock is s, then $\sigma = s\sqrt{\Delta}$.

3.4 Methodology to Directly Compare the Model to the Data

This section proposes a method to run simulations directly based on parameters estimated from the data for a particular pair of stocks. The values of the cross-correlation of the daily (weekly, or monthly) relative contribution of jumps can be compared to this same statistic calculated from many simulations based on the model in Section 3.3. Denote by $C_\times^{\text{data}}$ the observed value and by $C_\times^{\text{sim}}$ the simulated value of the moments, suppressing the time superscript (d), (w), or (m) for daily, weekly, and monthly intervals. Analyzing the quantity $C_\times^{\text{data}} - C_\times^{\text{sim}}$ will highlight disagreements between data and the model. The *a priori* expectation is not for this quantity to be identically zero. Rather, systematic deviations from the data may suggest directions for future investigation. Of course, running these comparisons necessitates parameter estimation from the observed price series for a particular pair of stocks. These parameters are estimated to ensure that the model is constrained to match certain moments well, such as the diffusive correlation, the diffusive volatility, the frequency of jumps, and the average magnitude of the jumps. The benefit of the parameterization presented above is that these moments correspond directly to individual parameters, and devising estimators is relatively straightforward. The remainder of this section discusses the estimation procedure of each parameter, and Table 4 summarizes the methodology.

Estimating parameters related to the diffusive component is straightforward. The diffusive volatility $\sigma^{(i)}$ can be estimated directly using the bipower volatility of each individual stock, averaged over the days in the sample. The diffusive correlation ρ_d is estimated through a statistic called bipower covariance. Although it is difficult to directly derive an estimator for bipower covariance in the same manner as bipower variation, the covariance can be backed out using the "portfolio method." If r_A and r_B are the five-minute returns of two stocks A and B, then the portfolio method considers the variance of an artificial portfolio whose five-minute returns are $\frac{1}{2}r_A + \frac{1}{2}r_B$. A simple manipulation shows that

$$\text{cov}\,(r_A, r_B) = 2 \cdot \text{var}\left(\frac{1}{2}r_A + \frac{1}{2}r_B\right) - \frac{1}{2}\,\text{var}(r_A) - \frac{1}{2}\,\text{var}(r_B). \tag{20}$$

The terms of the right-hand side of Equation (20) can be estimated jump-robustly from the bipower variances of the two stocks and the fictitious portfolio. Hence, the estimate for ρ_d is simply the bipower covariance calculated from Equation (20) divided by the product of the bipower volatilities of the two individual stocks.

Estimating parameters related to the jump processes involves tests designed to detect the presence of jumps. The frequency parameters $\lambda^{(i)}$ and λ_c are estimated using a Jiang-Oomen test at a 0.5 percent significance level. The estimate for $\lambda^{(i)}$ is the proportion of days in the sample on which a jump was detected for stock i. Similarly, the estimate for λ_c is the proportion of days in the sample in which a jump is detected for both stocks in the pair. The remaining parameters, however, require identification of the magnitudes of the observed jumps. To classify individual movements as jumps, this work follows a naive approach proposed by Andersen, Bollerslev, and Dobrev (2006).[7] Under the simplifying assumption that volatility is constant within the trading day, they show that a randomly drawn intraday diffusive return is approximately distributed $N(0, \Delta \cdot \sigma_t)$, where Δ is the sampling frequency and σ_t is the (true) diffusive volatility on day t. In the case of 5-minute intervals, as are used in this paper, Δ is set to 1/76. Thus, to detect intraday jumps, Andersen et al. propose

$$\kappa_j = r_{j,t} \cdot \mathbb{1}\left[|r_{j,t}| > \Phi_{1-\beta/2} \cdot \sqrt{\Delta \cdot BV_t}\right], \tag{21}$$

where β is the significance level of the test. As suggested by Andersen et al., β is set to $1 - (1-\alpha)^\Delta$ with $\alpha = 10^{-5}$. Intuitively, Equation (21) identifies as jumps any movements that are sufficiently large so as to be unlikely to be due to the diffusive process.[8]

Once these jumps are identified, estimating the remaining parameters is straightforward. A

---

[7]The primary concern of Andersen et al. (2006) was not to create a statistical test for jumps, and as such this method is not a common method in the literature to test whether an individual price movement is due to the jump process. Accordingly, I have not placed a description of this methodology under the Section 3.1. However, for the purposes of the model-to-data comparisons, this method provides a simple and intuitive method to conduct inference on the jump process.

[8]The test proposed by Barndorff-Nielsen and Shephard (2005) discussed in Section 3.1 provides an alternate method of estimating $\lambda^{(i)}$ and λ_c. Furthermore, $\lambda^{(i)}$ and λ_c can also be estimated by dividing the number of intervals identified as jumps or cojumps using the metholodolgy of Andersen et al. (2006) by the total number of intervals. These alternate methods of estimating these parameters do not change the estimates significantly, thereby providing an informal method of checking the robustness of the estimation methodology. As mentioned in Section 3.1, we use the estimates from the test by Jiang and Oomen (2008) since the test does not directly depend on $RJ_{BV,t}$.

cojump is identified when jumps are detected for both stocks in the same 5-minute interval. The sample correlation between the magnitudes of these cojumps is the estimate for ρ_j. The standard deviation for the magnitude of the jumps in stock i identified as part of the a cojump is the estimate for $\sigma_{cj}^{(i)}$. The sample standard deviation for the remaining jumps in stock i is the estimate for $\sigma_{\text{jump}}^{(i)}$. The method of identifying jumps also provides a manner of cross-checking the estimates for $\lambda^{(i)}$ and λ_c by computing the proportion of intervals in which jumps (or cojumps) are identified. In general, the two procedures return estimates that are comparable to each other, so the estimates derived from the Jiang-Oomen test are used.

4   Effects of the Underlying Parameters on C_{BV}

The Monte Carlo simulations described above provide insight into how the underlying properties of the process can affect the statistic C_{BV}, the cross correlation of the relative contribution of jumps (as measured using bipower variance) as defined in Equation (14). In this section, we systematically vary the underlying parameters of the model and study the effects on C_{BV}. We begin by varying the diffusive correlation ρ_d, fixing the frequency of cojumps λ_c, the magnitude (i.e., standard deviation) of the jumps σ_c, and the correlation ρ_c in the magnitude of the cojumps. We then vary the magnitude σ_c and the frequency λ_c of the cojumps. The final dimension altered is the correlation ρ_c in the magnitude of two components of the cojumps. For each parameter study, we summarize the results and present some intuition for the observations. The notation is summarized in Table 4, and the values used are tabulated in Table 5. Section 6 will place these theoretical results in context of the empirical findings.

Figure 5, which summarizes the effect of the diffusive correlation ρ_d on C_{BV}, shows the somewhat surprising observation that C_{BV} has a strong postive dependence on ρ_d. That is, even when jumps are small—volatility in a single jump is less than the daily volatility of the stock—, a large correlation in the underlying diffusive component can induce a residual correlation between the RJ_{BV} series and increase C_{BV} to a significantly positive number. Note, as will be discussed

in Section 7, 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 noticably positive for even the lowest value of λ_c considered. A reasonable explanation for this observation is that the statisic $RJ_{BV,t}$ erroneously treats large diffusive movements as "jump" movements during the calculation.[9] If the diffusive motion is correlated, stocks are more likely to have these large diffusive movements in the same day. The importance of this observation is that it suggests that factors entirely unrelated to the jump component have a nontrivial effect on C_{BV} in finite samples.

Another pattern evident from these simulations is that an increase in σ_c, which dials the magnitude of all jumps, or λ_c, which dials the the frequency of cojumps, increases C_{BV}. The intuition for both these observations is clearer. In the case that there is a jump within a day, the contribution to the total variation by this jump is likely to be larger if σ_c is larger. Potentially since σ_c dials the magnitude of the cojumps as well in this simplified setup, both $RJ_{BV,t}^{(1)}$ and $RJ_{BV,t}^{(2)}$ will be large on this day. Introducing ordered pairs where both elements are large may increase the correlation between the time series. A similar argument may apply for λ_c; as common jumps before more frequent, a greater number of days have large values of both $RJ_{BV,t}^{(1)}$ and $RJ_{BV,t}^{(2)}$.[10] For the sake of comparison, note that λ_c in the data is closer to the lower end of the range considered in these simulations.

The final dimension altered in these simulations is the correlation ρ_c of the magnitude of the two components of the cojumps, and the interesting observation from these studies is that the effect of ρ_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 ρ_d does. Figure 6 plots the result of the altering ρ_c, with ρ_d fixed at 0.40, a value comparable to that for two stocks in the same industry. As a potential explanation of this unexpected observation, consider the case when the magnitude σ_c of jumps is very large in comparison to the magnitude of diffusive movements. Then, $RJ_{BV,t}^{(i)}$

---

[9]Recall that bipower variation is robust to jumps since the effect of any large price movements on the computed variance is diluted by multiplying these large movements by smaller, neighboring movements. However, the effect of large diffusive movements are also erroneously diluted in this same manner.

[10]Of course, it is not true in general that introducing ordered paris where both elements are large increases the correlation. Adding these "large" elements will certainly increase the covariance of the two sets, but they will also increase the variance of the individual sets. The balance between these effects is difficult to determine *a priori*.

for $i = 1$ and 2 is already much larger for days with cojumps than for days without cojumps; the two components of the jumps are likely to be much larger than the diffusive movements even if the components are uncorrelated. In the case when σ_c is small and the correlation ρ_c is close to zero, however, it is possible to have "cojumps" that contain one large component that is identified as a jump and one small component that is not identified as a jump since it is comparable in magnitude to a diffusive movement. In this case, increasing ρ_c would increase the probability that large jumps in one stock are paired with large jumps in the other and would thus have a more significant effect on C_{BV}.[11] The simulations exhibit this behavior, as the curve for small σ_c slopes upward as ρ_c increases above 0.6. However, a simple mathematical model testing this conjectured explanation, such as the one developed in Appendix B, would be useful.

5   Data

5.1   Data Source and Cleanup

The data used in this paper come from a set of minute-by-minute prices for 22 commonly traded stocks in the S&P 100. They are obtained from the commerical data vendor `price-data.com`. These stocks are from three major industrial sectors—technology, finance, and food/agricultural— although no quantitative criteria were used to select these particular stocks. Each trading day contains price data from 9:35 AM to 4:00 PM. While trading on the New York Stock Exchange does start at 9:30 AM each day, the first five minutes of trading often involve a market reaction to overnight events. As a result, the behavior of the prices during this period may be substantially different from the rest of the intraday behavior, and this period is ignored in the analysis. By the same reasoning, this work ignores overnight returns entirely when calculating quantities for values that span multiple days; overnight returns also behave differently from intraday returns. Most stock data run from 1997 to 2010, athough data for certain securities are available for only a subset of

---

[11]Once again, we must note that increasing σ_c would also increase the variance of the relative contribution of jumps for a single stock. Thus, while this discussion can explain an increase in the covariance, the effect on the correlation is not clear *a priori*.

that time period. For example, data for Google (GOOG) is of course only available after its IPO in 2004, and data for Exxon-Mobil (XOM) is available only after the Exxon/Mobil merger in 1999. Table 1 lists the stocks used in this analysis along with the industries, the start dates, the end dates, and number of trading days observed.

The stock data also require some cleaning. First, the data are adjusted backward for stock splits, using data from Yahoo! Finance as a guide to distinguish stock splits from large jumps and other economic events. Of course, stock splits are also distinguishable in the data since the stock price falls by exactly one-half or two-thirds. Days with data entry errors, such as missing price values, are very rare but are nonetheless excluded from the analysis. A number of trading days were missing for certain stocks. To compensate for this lack of data, when calculating statistics that rely on data from two stocks, the time series are aligned properly: any date that is missing from one series is discarded from the other.

5.2 Microstructure Noise

An important aspect of financial econometrics that has thus far been ignored is that the data are a noisy and incomplete representation of the true price process. First, the econometrician only observes prices when the security is traded, which occurs at discrete time intervals. Second, instead of observing the efficient price $p(t)$, the econometrician observes the price $p^*(t)$, which is contaminated with "microstructure noise." That is, $p^*(t) = p(t) + \epsilon(t)$, where $\epsilon(t)$ is a mean-zero noise term often attributed to market frictions that cause a short-run deviation between the efficient price and the price at which the security is traded. One major source of microstructure noise is the bid-ask spread, which refers to the simple observation that the price observed is either a buying price or a selling price; for example, a trader wishing to buy a certain number of shares of an equity immediately would put in a "bid" order at a price that is likely slightly higher than the efficient price. In the domain of very high frequency data, the observed prices can bounce between "bid" prices and "ask" prices, and this bid-ask bounce may introduce spurious volatility into the estimates. Another source of microstructure noise is discretization error. Stock prices are quoted in cents whereas the

efficient price (as seen from Equation (2)) can be any positive real number. Of course, discretization errors were more important before decimalization of the stock market around 2001, when prices were quoted in sixteenths of a dollar. Accounting for microstructure noise is an important part of financial econometrics.

A simple method to minimize the effect of microstructure noise involves choosing the sampling interval $\Delta \equiv 1/M$ in a careful manner. Under idealized circumstances, asymptotic consistency of the estimators would suggest that Δ ought to be taken to zero so that the prices are sampled as frequently as possible. However, microstructure noise can contaminate these estimates. In fact, in the case where $\epsilon(t)$ is independent (both serially and with respect to all other quantities in the price process) and identically distributed, a simple calculation shows that the estimated realized variance will increase without bound as the sampling interval decreases. A natural solution would be to use a subsample of all available data by increasing the sampling interval above the minimum possible value. Plotting the unconditional realized variance as a function of the sampling interval usually exhibits the pattern that the estimated variance increases for small (1-, 2-, or even 5-minute) intervals and approaches a flat, constant value for intervals that are sufficiently long. Andersen, Bollerslev, Diebold, and Labys (2000) propose such a "volatility signature plot" to determine a sufficiently short interval that does not increase the unconditional volatility much above the value when using long intervals. Using a volatility signature plot, this paper settles on using 5-minute intervals to calculate the estimates. Monte Carlo evidence from Huang and Tauchen (2005), who consider a model that includes microstructure noise, suggests that using staggered returns improves the performance of jump tests. Unfortunately, this coarse sampling procedure necessitates throwing away 80 percent of the data. While other methods have been developed to avoid this unfortunate problem, this work does not utilize any such methods.

6 Empirical Findings

The primary question this paper addresses is whether predictions from a standard model of stock prices agree with a particular moment in the data when considering pairs of stocks: the cross-correlation $C_\times$ of the relative contribution of jumps to total price variation $RJ_{\times,t}$. A previous section studied some of the properties of the theoretical model we wish to study, and this section studies empirical regularities of the statistics we are considering. To accomplish this goal, this section analyzes the data described in Section 5 and qualitatively summarizes the most salient observations. The analysis emphasizes comparing pairs of stocks from the same industry (often referred to as "within-industry pairs") to pairs of stocks from different industries ("out-of-industry pairs"). We then connect these empirical observations to the theoretical findings of the previous section and provide guidance as to how to explain some of the results presented in this section. Finally, Section 7 asks the quantitative question that is at the heart of this paper. In that section, the data are used to calibrate model parameters directly to specific pairs of stocks; the data and model moments are then compared. The empirical observations made in the current section inform the statistical comparisons made in this final section.

6.1 Relative Contribution of Jumps for a Single Stock

For any jump-robust estimator $\times$ of quadratic variation, the statistic $RJ_{\times,t}$ is a noisy estimator of the relative contribution of jumps to the total variance. Figure 1 considers the jump-robust estimator BV_t and plots the time series of $RJ_{BV,t}$ for KO for each day on which the vendor provided data. An important observation is that the variation in the statistic is so large that 27.4 percent of days in the sample have $RJ_{BV,t} < 0$, even though the statistic is a consistent estimator for a nonnegative quantity. Similar observations hold for $RJ_{MinV,t}$ and $RJ_{MedV,t}$, although plots are not included in this paper. Indeed, the daily variation in this statistic overshadows any possible time trend in the plot.

    Summarizing the statistic using yearly averages uncovers some noticable patterns, partly by

mitigating the effect of the large daily variation. Table 2 tabulates the mean of the RJ_{BV} series for various stocks by year, from 1997 to 2010, also listing the yearly means and standard errors when grouping the stocks by industry. This table indicates that some stocks do show a significant time trend. For example, AAPL, BK, and HPQ—among others—have mean values of RJ_{BV} which are signficantly higher in the earlier portion (1997 to 2001) of the sample than in the latter portion (at the 5 percent 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 thereafter. This trend is noticeable when averaging across stocks as well, as evident from the final four lines of the table; however, large within-industry variations induce larger standard errors in these averaged quantities. The data in Table 2, coupled with other statistical analyses, suggest that the statement that $RJ_{\times,t}$ is similar across stocks and across time is not entirely accurate, although the mean value of $RJ_{\times,t}$ seems to follow a similar pattern for most stocks.

6.2 Relative Contribution of Jumps for Pairs of Stocks

Similar exploratory studies on bivariate series of $RJ_{BV,t}$ reveal regularities about the relation between the cross-correlations of the daily, weekly, and monthly relative contributions of jumps (denoted $C_{BV}^{(d)}$, $C_{BV}^{(w)}$, and $C_{BV}^{(m)}$, respectively). In particular, there are interesting patterns when comparing pairs of stocks within the same industry to pairs from different industries. Figure 2 graphically summarizes these three quantities for all other stocks with BAC (a finance company) and DELL (a tech company); the data points are color-coded as described in the caption. These stocks were chosen since they exemplify two different patterns. The plot for BAC indicates 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. In the plot for DELL, the industry pattern is somewhat less clear. To quantify this relationship and summarize it across all the stocks in the set, Table 3 tabulates the average C_{BV} by industry. For many stocks—especially those in the food and finance industries—the results suggest that within-industry correlations are larger than out-of-industry correlations, although the difference is not statistically signficant if using a simple

two-sample t-test. As a result, this finding can be viewed as guiding principle that "similar" stocks tend to have higher C_{BV} values.

The data exhibit a sort of temporal pattern as well in that for a particular pair of stocks, the cross-correlation calculated at the monthly level is larger than that at the weekly level, which is in turn larger than that at the daily level. That is, when moving from $C_{\times}^{(d)}$ to $C_{\times}^{(w)}$ to $C_{\times}^{(m)}$, the correlation grows stronger. Figure 2 motivates this observation to some extent in that the points cluster closer to the right when moving from the "daily" line to the "monthly" line. Note that this behavior is evident when analyzing the numbers in Table 3; the mean values of $C_{\times}$ tend to increase for most stocks when moving from daily to weekly to monthly periods. Figure 3 displays this information graphically for the case where the jump-robust estimator is BV by plotting $C_{BV}^{(m)}$ and $C_{BV}^{(w)}$ against $C_{BV}^{(d)}$. The figure partitions the set of all pairs based on the industry grouping. Note that the slope of the regression lines through each individual scatter plot is a rough proxy for the effect. Figure 3 clearly suggests that the temporal pattern is evident for both within-industry and out-of-industry pairs, although it is slightly more pronounced for within-industry pairs. In the rest of this section as well as the subsequent section, we attempt to study this temporal pattern along with the industry clustering described above in the context of the model presented in Section 3.3.

A possible hypothesis for these observations is that the higher correlation within industries is due to the fact that jumps for similar stocks tend to occur on the same day, possibly due to industry-relevant economic news released on a given day that affects multiple stocks. One way to test this hypothesis is to study the relation between $C_{\times}$ and JC, the correlation in the indicator variables for detecting the presence of a jump on a given day. A positive correlation would suggest that the patterns observed in $C_{\times}$ may be related to lack of independence in arrival times of the underlying jump process; the mere presence of a common jump component can induce both high JC and high $C_{\times}$. It must be noted that the magnitude of JC, unlike the value of $C_{\times}$, does not take into account quantities such as the size of the jumps or the correlation of their magnitudes. A high correlation between JC and $C_{\times}$ would suggest that these additional aspects are not important factors in driving $C_{\times}$. Figure 4 plots example comparisons between $C_{MinV}^{(d)}$ and JC at the 1 percent

and 0.1 percent signficance levels. While the figure for BAC shows a similar industry ordering between JC and $C_{MinV}^{(d)}$,[12] 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 percent significance level, against $\tau_{BV}^{(d)}$ for all 231 pairs of stocks. That is, each data point in the figure corresponds to a single pair of stocks. The slope of a regression line through this scatter plot—as can be anticipated from the diagram—is insignificant at any reasonable significance level. A reasonable conclusion is that the statistic C_{BV} (and all related statistics) contains information that cannot be explained purely by tests for the presence of jumps, and quantities such as size and correlation in magnitudes are also important.

These parameter studies presented in Section 4 partially explain some of the observations presented above. First, the industry pattern may be due to something as simple as a misleading artifact of a large correlation in the diffusive movements in the stocks in the same industry; the simulations above showed that there may be a considerable (finite sample) effect on C_{BV} stemming from this correlation that has nothing to do with the jump component of these price series. Another potential hypothesis for the industry pattern may have been that the frequency of the cojump process for stocks from the same industry is larger; as shown in the parameter studies, a larger frequency of cojumps would have increased C_{BV}. However, a large cojump frequency would also have increased JC considerably, which is not observed empirically. In fact, the observation that there is little association between JC and C_{BV} may be partially explained by the theoretical result that C_{BV} is strongly affected by the magnitude of the jumps in the process. It is also reasonable that stocks from the same industry may have a larger correlation in the magnitudes of their cojumps, but the parameter studies suggest that the effect of this correlation on C_{BV} would be minimal and almost nonexistent if the magnitude of the cojumps is sufficiently large.

---

[12]Throughout this paragraph, alternate statistics such as C_{MinV} and $\tau_\times$ have been used to offer an alternative to the standard ones we consider. All these statements hold true when using C_{BV} or any other statistic.

7 Implications Relative to the Model

7.1 General Observations

Since systematic parameter studies over a large parameter space are not feasible, a direct method of comparing the model to empirical data is to calibrate the parameters to a particular pair of stocks. Direct comparisons between the model and observation—to test the effectiveness in matching the cross-correlation of the relative contribution of jumps—are straightforward if using the method described in Section 3.4. Table 6 lists the results from two sample comparisons; it shows both C_{BV}^{sim}, which is computed from simulations with parameters calibrated to the particular pair of stocks, and C_{BV}^{data}, which is estimated from the data. 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 Monte Carlo setup described in Sections 3.3 and 3.4, the true data is simply one sample path. In Table 6a, the mean and standard deviations of the simulated distributions of C_{BV}, C_{MinV}, and C_{MedV} are listed for the finance/finance pair BAC/C. The BAC/C pair is an example of the common "undershoot" error observed in many pairs of stocks. That is, the statistics from the observed data are larger than those from the simulations and $C_{\times}^{\text{data}} - C_{\times}^{\text{sim}}$ is consistently positive. Furthermore, this difference grows when moving from daily to monthly periods. However, a number of stock pairs show close agreement between empirical and simulated data. For example, Table 6b shows that the empirical data for the finance/food pair BK/KFT lie within one standard deviation of the simulated means for all listed statistics. Perhaps more importantly, the quantity $C_{\times}^{\text{data}} - C_{\times}^{\text{sim}}$ does not possess a systematic sign. These two examples indicate that the model agrees with the observed values in some cases but not in others.

It is possible to summarize these observations by grouping pairs of stocks based on industries, and we see that within-industry pairs tend to have a poorer match (on average) between the observed values of C_{BV} and the model predictions. Table 7 averages both C_{BV}^{sim} and C_{BV}^{data} over all pairs of stocks within a particular industry grouping. This table first reinforces the observation that the increase in C_{BV} from daily to monthly levels is slightly more pronounced for within-industry

comparisons, especially within the finance and food industries; however, this increase is still noticable in out-of-industry comparisons, even though it is not as strong. The interesting observation seen in Table 7a is that C_{BV}^{sim} differs noticeably from C_{BV}^{data} at all levels. The finance and food industries exhibit a clear example of the "undershoot" error described above, and the magnitude of the error is relatively stark. The tech industry exhibits the opposite error in that the simulated values from the model tend to be larger than those from the data. When considering out-of-industry comparisons, the model does not match the data when considering finance/food combinations, but it does match both finance/tech and food/tech pairs well on average.

Figure 7 graphically highlights the observations made earlier by plotting the model predictions against the observed values for the data. Perfect prediction would correspond to all the data points lying on the $45°$ line. Figure 7a considers within-industry pairs, and the general observation is that the datapoints lie away from the $45°$ line. However, large deviations occur in *both* directions, suggesting that the model does not systematically differ from the data in any particular manner. Figure 7b, which considers out-of-industry pairs, shows that the points lie much closer to the $45°$ line. Indeed, the observed values of C_{BV} lie less than one standard deviation from the model predictions for most out-of-industry pairs.

7.2   Summary Statistics

A summary statistic for discrepancies between simulation and data can formalize the observations made in the previous paragraphs. The quantity $\left| C_{BV}^{\text{sim}} - C_{BV}^{\text{data}} \right|$ is a measure of the deviation between the simulated mean and the observed statistic. Table 8 tabulates this quantity when averaged over different industry pairs and highlights a few observations.[13] First, the discrepancy is larger for within-industry pairs than out-of-industry pairs. Second, the increase in this discrepancy when moving from daily to monthly periods is more significant for within-industry pairs than out-of-industry pairs. Note that this quantity penalizes discrepancies between the observed value and

---

[13]It must be noted that the quantities in Table 8 are not simply the absolute values of the difference between C_{BV}^{sim} and C_{BV}^{data} in Table 7. To calculate the values in Table 8, the absolute value is taken before the average.

the model-generated (simulated) value in both directions. Thus, the fact that the increase in the discrepancy in the tech/tech pairing is larger than would be anticipated from Table 7a is a byproduct of tech/tech pairs having large deviations in both directions whereas finance/finance and food/food pairs only had large "undershoot" errors.

A similar statistic involves normalizing the difference between the observed value of C_{BV} and the prediction from the model by the standard deviation of the distribution of the simulated values. For any particular parameter choice, the distribution of C_{BV} in the simulations is approximately normal, as tested using a Kolmogorov-Smirnov test at a 1 percent significance level. Using this test as justification for assuming that the true distribution of C_{BV} in the simulations is exactly normal, the statistic $\left(C_{BV}^{\text{sim}} - C_{BV}^{\text{data}}\right)^2 / \sigma^2_{C_{BV}^{\text{sim}}}$, where $\sigma^2_{C_{BV}^{\text{sim}}}$ is the variance of the simulated distribution, is distributed χ^2_1. It is difficult to aggregate these statistics over a particular group of observations—such as $C_{BV}^{(d)}$ for all finance/finance pairs—since the individual χ^2_1 statistics are not from independent normal distributions: the same stock is of course part of many different pairs. As a results, the sum of the test statistic over k pairs is in general not distributed χ^2_k.[14] However, as a heuristic way to judge the goodness-of-fit of the model predictions to the data, the average value of the χ^2_1 statistic can be calculated across all pairs in the group. Table 9 lists these average statistics when considering groups in a particular industry pair when the statistic is calculated over a particular time interval: these statistics are then aggregated over larger groups and finally across all within-industry and out-of-industry pairs in the same manner. One must be careful about interpreting these statistics, as the distribution of the average is also unknown due to the correlation. However, these averages can be roughly used to compared fits between groups, even if formal significance tests are not currently possible.

Table 8 and Table 9 highlight perhaps the most central finding in this paper: the model presented in Sections 3.3 and 3.4 often cannot replicate the observed values of the cross-correlation of the

---

[14]It is, of course, possible to estimate the covariance matrix C_{BV}^{sim} for various a collection of different pairs of stocks. Recall that if X is an $n \times 1$ normal with mean 0 and covariance matrix Σ, then $X'\Sigma^{-1}X$ is distributed χ^2_n. This relation between the normal and the chi-squared that can be used to develop a goodness-of-fit statistics that is indeed distributed chi-squared. However, joint estimation of the covariance matrix Σ involves simulating many price series simultaneously, which is computationally expensive.

relative contribution of jumps in the data, even when parameters are calibrated to particular pairs of stocks. Moreover, Table 8 suggests that this discrepancy is systematic in the sense that out-of-industry predictions are on average more accurate than within-industry predictions. However, Table 7 shows that these discrepancies cannot be attributed simply to something like the failure to replicate the temporal pattern; this is because the temporal pattern exists in certain out-of-industry pairs as well, and the model predictions tend to "overshoot" the temporal pattern in the tech/tech case. Figure 7 reinforces this observation by showing that many out-of-industry pairs show a small discrepancy between the model prediction and the observed data while within-industry pairs tend to deviate considerably in both directions.

Table 9 takes into account the standard deviations of the simulated distribution in its statistics, and the implications are similar. First, even when accounting for the standard deviation of the distribution of simulated values of C_{BV}, within-industry comparisons are slightly worse than out-of-industry comparisons. The finance/food group, however, seems to show a strong discrepancy between model and data, as was also noted above. The second major implication is that the discrepancy between model and data does *not* increase as the time interval considered increases, if we correct for the standard deviations of the simulated distributions. This result is due to the fact that the standard deviations of the simulated distributions increase as the time interval increases. These results suggest that while the simulations are somewhat accurate in predicting the observed value of C_{BV} for most out-of-industry pairs, they may be missing some systematic effect that relates stocks in the same industry. Determining this missing factor would be an interesting extension to this work.

8 Discussion

This paper presents a first step towards cross-stock comparison of the relative contribution of jumps to total price variance. While the empirical literature has studied the relative contribution of jumps to total price variance for a single stock, it has not considered the natural question of the relationship

of this quantity between pairs of stocks. This paper focuses on the cross-correlation of the time series of the relative contribution of jumps, denoted C_{BV}. By calculating the relative contribution of jumps at the daily, weekly, and monthly levels, it computes three measures of the cross-correlation, one at each level. First, we find that the cross-correlation of the relative contribution of jumps tends to increase when moving from daily to weekly to monthly time periods. Secondly, the cross-correlation is larger for pairs from the same industry than for pairs from different industries. The results presented in this paper suggest that the magnitudes of the jumps—and the correlations in the magnitudes of cojumps—likely play a significant role in these findings. This observation stems from studying correlations between simple binary time series for the *presence* of jumps, which are of course only sensitive to the frequency of the jump process. These correlations do not show the same patterns as do time series of the relative contribution of jumps to total price variance.

To help explain these findings, this paper considers a simple jump-diffusion model for the price processes of a pair of stocks. Due to the complexity of the statistic considered in this paper, it is difficult to make analytic progress even when using a simple model such as the one considered in this paper. As a result, a series of Monte Carlo experiments were conducted to study the effect of the underlying parameters. Somewhat surprisingly, these experiments suggest that factors entirely unrelated to the jump processes can affect the cross-correlation between the relative contribution of jumps. For example, a high diffusive correlation between the two price processes can cause the (estimated) relative contribution of jumps to become correlated. Furthermore, certain parameters closely tied to the jump process—such as the correlation between the magnitudes of cojumps— have negligible effects on the cross-correlation considered. Section 4 provides some discussion of potential reasons for these observations. The most interesting finding of this study is that when this simple jump-diffusion model is calibrated to parameters estimated from the pairs of stocks in the data, it is only somewhat accurate in predicting the observed value of the cross-correlation. The model predictions still replicate the increase in the cross-correlation from daily to weekly to monthly time windows, but the magnitudes of the cross-correlation are usually underpredicted. The discrepancy between model predictions and the data is larger for pairs of stocks from the same

industry than for pairs of stocks from different industries. This systematic discrepancy suggests that the model is missing a factor that relates stocks that are "similar," such as those from the same industry.

This paper also presents some basic empirical findings related to the relative contribution of jumps for a single time series, and one of the main observations from those studies is that the relative contribution of jumps may have a significant time trend. Some of the discrepancies between the model predictions and the empirical observations may be related to the fact that underlying parameters of the model may also have a nonnegligible time dependence in reality. For example, a possible explanation for the reason that the discrepancy between model and data is larger for within-industry pairs is that the frequency of jumps (even idiosyncratic jumps) vary over time. If it is the case that there is a correlation between the underlying processes of the idiosyncratic jump frequency for two stocks in the same industry, then C_{BV} would be larger than would be predicted from a model that does not include such a time-varying frequency. Another explanation may relate to the fact that the jump intensities themselves are time varying; there is some evidence for this phenomenon (Chan & Maheu, 2002; Maheu & McCurdy, 2004). Introducing time-varying jump intensities into the model would again increase the cross-correlation, especially at the weekly and monthly levels. A simple future study to control for time-varying parameters might involve partitioning the observed sample into smaller subsamples to limit the degree to which the parameters might vary over the sample.[15]

The work in this paper suggests a number of potential directions for future study, and we will briefly outline some ideas in this paragraph. First, if the true distribution of the jump magnitudes is a mean-zero normal, then it is possible for jumps to have small magnitude. However, the Jiang-Oomen test will not classify such small price movements as jumps, and this fact will bias the estimators.

---

[15]It should be noted that this proposed explanation does not necessarily contradict the observation that the discrepancy (normalized by the standard error) between the model and the data does not increase as the time interval increases. This proposed explanation instead is concerned with variation in the parameters over significantly longer intervals, such as years. Furthermore, the author has conducted some model-to-data comparisons over shorter time intervals, separating the entire sample into three equal intervals. Preliminary results are similar to the ones presented in this thesis, and as such, they are not reported here. However, a more careful method to determine appropriate intervals in an effort to limit the effects of large movements of the underlying parameters over time may be useful.

Second, the method used by Andersen et al. (2006) to classify each individual price movement as either a diffusive movement or a jump neglects the intraday pattern in volatility. Since volatility within a trading day is large near the start and finish of the trading day and lower near the middle, the method adopted from Andersen et al. will underestimate jumps near the middle of the day and overestimate jumps near the beginning and the end. Bollerslev, Todorov, and Li (2011) develop an estimator to classify individual movements as jumps that accounts for this phenomenon.

The model itself is also especially simple, and more sophisticated approaches may help explain some of the discrepancies between the model and data. Many recent papers (Bollerslev et al., 2011; Huang & Tauchen, 2005) use a stochastic volatility model for the diffusive volatility $\sigma(t)$ instead of the unrealistic simplifying assumption that volatility is constant in time. Furthermore, the specification in this paper of jumps being drawn from a normal distribution may be modified: Bollerslev et al. (2011) draws jumps from a truncated normal to ensure that the magnitude of jumps is larger than the average magnitude of a diffusive movement. Introducing microstructure noise into the model may be interesting, following Bollerslev et al. (2011). While the calibrated parameters have often been able to replicate the increase in the cross-correlation statistic from the daily to the monthly levels, it may also be the case that this increase is also partly due to the fact that microstructure noise averages out at these longer time intervals.

The appendices present initial work to address some of the aforementioned ideas. Appendix A considers a simplified model and computes the bias of the estimators used in this paper, assuming that the true distribution of the jumps is mean-zero normal. It also briefly discusses some benefits of using a truncated normal distribution for the jumps, an idea proposed above. Appendix B attempts to analytically characterize some of the quantities considered in this paper. It explores a simplification that allows for an explicit computation of the distribution of the relative contribution of jumps to total price variance and discusses possible methods to compute properties of the cross-correlation of the relative contribution of jumps.

9 Conclusion

While the relative contribution of jumps to total price variance had been studied in the literature, this paper is the first to conduct systematic comparisons of this quantity across pairs of stocks. The statistic considered in most detail in the cross-correlation of the relative contribution of jumps. We find that this cross-correlation is strongly affected by the time interval over which the relative contribution of jumps is calculated: calculating the relative contribution of jumps over longer intervals tends to increase this cross-correlation. Furthermore, the cross-correlation is larger for pairs of stock within the same industry than for pairs of stock from different industries. Another main finding in this paper is that a standard model for the price processes of stocks, when calibrated to particular pairs of stocks in the data, cannot replicate the empirical values of the cross-correlation of the relative contribution of jumps. In fact, the model predictions differ from the empirical observations systematically in that predictions for out-of-industry pairs are on average more accurate than those for within-industry pairs. Much work is needed to explain the systematic discrepancies between the model predictions and the observed values of the statistic in the data. While some possible explanations have been presented above, other explanations are of course possible, and additional work is needed to determine which ones match the observed data well. Indeed, this study can be viewed in a different angle: it may be possible to use the effectiveness of a model to match the cross-correlation moment in the data as a method of evaluating the effectivess of the model in describing the price processes of a pair of stocks.

A Complications with Using a Gaussian Model for Jumps

An assumption of the model presented in Section 3.3 is that the magnitude of the cojumps is jointly normal (Gaussian) with mean zero. If this assumption were in fact true, then the price process would include a number of small "jumps," with magnitude close to zero.[16] Such movements should

---

[16]A normal distribution gives positive probability to any open interval, so one may also object that a normal model for jumps is unrealistic since arbitrarily large jumps occur with positive probability. The concern in the setup is that if the normal distribution is mean zero, then jumps comparable in magnitude to a diffusive movement—that is, realizations

likely not be characterized as jumps and would certainly not be distinguished from the diffusive price process using jump-detection tests. As a result, such an assumption maybe unrealistic.

Despite this unrealistic assumption, this appendix considers the case where the true jump process is indeed mean zero normal. In this setup, the estimation procedure described in Section 3.4 is biased, and this appendix calculates the sign and magnitude of the bias. The Jiang-Oomen test will not identify days containing small jumps as "jump days," and our estimates for λ_i and λ_c will be biased downward. If we employ the procedure proposed by Andersen et al. (2006), then we will identify only the large jumps as jumps. As a simple model of this phenomenon, consider a situation in which the a sequence X_i is drawn from a standard normal distribution but that the econometrician only observes X_i if $|X_i| > c$ for some cutoff value c. The econometrician does not know of this censoring and uses the sample standard deviation of the observed X_i to estimate the standard deviation of the underlying data generating process. This estimate will clearly be larger than 1. Thus, the estimates for $\sigma_{cj}^{(i)}$ and $\sigma_{\text{jump}}^{(i)}$ are biased upward.

Finally, this same model can be used to predict the sign of the bias in the estimator for ρ_c. Suppose (X_i, Y_i) is drawn from a multivariate normal where the marginals are standard normal and the underlying correlation is ρ. Suppose further that the econometrician only observed (X_i, Y_i) for $|X_i| > c$ and $|Y_i| > c$. The sample correlation will overestimate the magnitude of ρ. The estimate will converge to

$$\frac{\mathbb{E}[XY \mid |X| > c \cap |Y| > c]}{\text{sd}(X \mid |X| > c) \cdot \text{sd}(Y \mid |Y| > c)}. \tag{22}$$

Figure 8 shows the quantity in Equation (22) for various values of the cutoff c and the true underlying correlation ρ. The values in Figure 8 are calculated by numerically evaluating the integral corresponding to Equation (22). The bias becomes more severe as the cutoff increases, as would be anticipated. Furthermore, as is clear from symmetry arguments, the estimator is unbiased when the true correlation is -1, 0, or 1.

---

approximately in the neighborhood $(-\sigma_d, \sigma_d)$, where σ_d is the standard deviation of a diffusive movement—happen with nontrivial probability.

B A Simplified Model to Study Effects of Parameters

This section presents a simplified model to analytically study the relative contribution of jumps to total price variance. It also discusses a method to compute the relative contribution of jumps to total price volatility in this simplified model, although it will be clear that analytic progress is difficult. One goal of this model is to abstract away from statistical considerations. In the Monte Carlo experiments of Section 4, two forces are at work: (1) the actual dependence of the cross-correlation of the relative contribution of jumps to total price variation on the underlying parameters of the model, and (2) statistical artifacts introduced by using C_{BV} as our statistic for this quantity. While it is possible in these Monte Carlo simulations to abstract away from the second consideration (as the true price process is known), it would be useful to have analytic expressions for the relative contribution of jumps.

We present the model for the price process. Consider two stocks labeled A and B and suppose that each stock undergoes a series of N price movements in a single day. Denote these price movements $\{(r_{A,i}, r_{B,i})\}_{i=1}^{N}$. A particular price movement i is *diffusive* if $(r_{A,i}, r_{B,i}) \sim N(0, \Sigma)$, where Σ is a covariance matrix such that each price movement has marginal variance 1 and the two have correlation ρ_d.[17] A price movement i is called a *jump* if $(r_{A,i}, r_{B,i}) \sim N(0, \Sigma_J)$, where

$$\Sigma_J = \begin{pmatrix} \sigma_c^2 & \rho_c \sigma_c^2 \\ \rho_c \sigma_c^2 & \sigma_c^2 \end{pmatrix}.$$

Of course, for the setup to be economically sensible, it should be the case that $\sigma_c > 1$, although nothing in the calculations require it. Let J be an indicator variable that takes the value 1 with probability p and 0 with probability $1 - p$ independently of everything else in this setup. If $J = 0$, then a sequence of price movements within a day is jump-free; that is, $(r_{A,i}, r_{B,i})$ are diffusive for all i from 1 through N. If $J = 1$, then the day contains a jump; that is, $(r_{A,i}, r_{B,i})$ are diffusive for i from 1 through $N - 1$, and $(r_{A,N}, r_{B,N})$ is a jump. Note that this model only allows for cojumps;

---

[17]Note that setting the variance of a diffusive movement to 1 is purely for normalization purposes that makes some of the subsequent calculations less tedious.

there is no mechanism for idiosyncratic jumps.

This model then allows for calculations of the relative contribution of jumps. Denote by R_A the relative contribution of jumps of stock A. Let

$$X_A = \frac{r_{A,N}^2}{\sum_{j=1}^{N-1} r_{A,j}^2 + r_{A,N}^2}. \tag{23}$$

If a day is jump-free, then the relative contribution of jumps is 0. Otherwise, the relative contribution of jumps is X_A. Therefore,

$$R_A = \begin{cases} 0 & \text{if } J = 0 \\ X_A & \text{if } J = 1 \end{cases}. \tag{24}$$

That is, $R_A = X_A J$. Note that $R_B = X_B J$ and due to the symmetry of the problem, the distributions of R_A and R_B are the same. Then, the cross-correlation of the relative contribution of jumps to total price variance is given by

$$\text{corr}(R_A, R_B) = \frac{\text{cov}(R_A, R_B)}{\sqrt{\text{var}(R_A)\,\text{var}(R_B)}} = \frac{\text{cov}(R_A, R_B)}{\text{var}(R_A)}. \tag{25}$$

We have that $\text{var}(R_A) = \mathbb{E}[R_A^2] - \mathbb{E}[R_A]^2$. But, $R_A = X_A J$, and since X_A is independent of J, $\mathbb{E}[X_A J] = \mathbb{E}[X_A]\mathbb{E}[J] = p\mathbb{E}[X_A]$. Since $J^2 = J$, we have that $\text{var}(R_A) = p\mathbb{E}[X_A^2] - p^2\mathbb{E}[X_A]^2$. Similarly,

$$\text{cov}(R_A, R_B) = \mathbb{E}[R_A R_B] - \mathbb{E}R_A \mathbb{E}R_B = \mathbb{E}[X_A X_B J^2] - (\mathbb{E}[X_A J])^2 = p\mathbb{E}[X_A X_B] - p^2\mathbb{E}[X_A]^2.$$

Thus,

$$\text{corr}(R_1, R_2) = \frac{\mathbb{E}[X_A X_B] - p\mathbb{E}[X_A]^2}{\mathbb{E}[X_A^2] - p\mathbb{E}[X_A]^2}. \tag{26}$$

Note that the dependence on the frequency of cojumps comes entirely from p; a smaller p corresponds to a smaller frequency. Equation (26) shows the ambiguity in the effect of the frequency of cojumps discussed in a footnote in Section 4.

The remainder of this appendix is devoted to computing the terms in Equation (26). It will be possible to characterize the distribution of R_A (by characterizing the distribution of X_A), but we have not found a method of computing $\mathbb{E}[X_A X_B]$. We will first compute the density of X_A. Note that since $r_{A,N} \sim N(0, \sigma_c^2$, we have that $r_{A,N}^2 \sim \text{Gamma}(1/2, 2\sigma_c^2)$. Furthermore, $\sum_{i=1}^{N-1} r_{A,i}^2$ is a sum of squares of $N-1$ independent standard normals; thus, it is distributed χ_{n-1}^2, which is also $\text{Gamma}((n-1)/2, 2)$. It can be shown that if $S \sim \text{Gamma}(m, \theta_S)$ and $T \sim \text{Gamma}(t, \theta_T)$ then $S/(S+T) \sim G(R)$ where $G(r) = \theta_S r / [\theta_S r + \theta_T (1 - r)]$ and $R \sim \text{Beta}(a, b)$. A simple transformation of the density function of a beta random variable shows that the density of $G(R)$

$$f_{G(R)}(t) \propto \frac{t^{a-1}(1-t)^{b-1}}{\theta_A(1-t) + \theta_B t}. \tag{27}$$

Equation (27) is a first attempt at characterizing the distribution of the relative contribution of jumps to total price variance for a single stock, something that has not been done explicitly in the literature before. While this equation applies to a simplified model, the techniques used may apply to more general models.

Equation (27) can be used to find the first and second moments of X_A, noting that $a = 1/2$, $b = (N-1)/2$, $\theta_A = 2\sigma_c^2$ and $\theta_B = 2$. Using a table of integrals, we find that

$$\mathbb{E}X_1 = \frac{1}{2\sigma_c} \Gamma\left(\frac{N}{2}\right) {}_2\tilde{F}_1\left(\frac{3}{2}, \frac{N}{2}, \frac{N}{2} + 1, 1 - \frac{1}{\sigma_c^2}\right) \quad \text{and} \tag{28}$$

$$\mathbb{E}[X_1^2] = \frac{3}{4\sigma_c} \Gamma\left(\frac{N}{2}\right) {}_2\tilde{F}_1\left(\frac{5}{2}, \frac{N}{2}, \frac{N}{2} + 2, 1 - \frac{1}{\sigma_c^2}\right), \tag{29}$$

where ${}_2\tilde{F}_1(a, b, c, z) = {}_2F_1(a, b, c, z)/\Gamma(b)$, $\Gamma(\cdot)$ is the gamma function, and ${}_2F_1$ is the Gaussian hypergeometric function. The computation given above for $\text{var}(R_A)$ can be used to determine the variance of the distribution of the relative contribution of jumps for a single stock.

Computing the correlation between the relative contribution of jumps for two stocks involves computing $\mathbb{E}[X_A X_B]$, which is very complicated. Work in this setting suggests that expecting a clean formula for the joint distribution of (X_A, X_B) may be unrealistic. Moreover, even computing

this moment explicitly may be difficult. As can be seen, this computation would involve a product of χ^2 random variables that are derived from correlated normal random variables. Royen (1990) gives an expression for the joint distribution of such random variables, and it is an infinite series in terms of the Γ function and is difficult to manipulate. As a result, we do not pursue an explicit computation of the correlation of the relative contribution of jumps any further.

C Tables

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

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 RJ_{BV}, calculated by year and by stock. The $RJ_{BV,t}$ time series is calculated at the daily level using 5-minute increments.

Stock	Overall	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
AAPL	0.0763 (0.0022)	0.137 (0.0096)	0.138 (0.0084)	0.121 (0.0087)	0.0808 (0.0081)	0.0634 (0.0078)	0.0793 (0.0075)	0.0923 (0.0075)	0.0634 (0.008)	0.061 (0.0078)	0.0424 (0.0068)	0.0446 (0.0078)	0.0383 (0.0071)	0.062 (0.0069)	0.0635 (0.008)
BAC	0.0632 (0.002)	0.118 (0.0087)	0.0878 (0.0073)	0.0875 (0.0074)	0.119 (0.0078)	0.0679 (0.0081)	0.0327 (0.0066)	0.0352 (0.0074)	0.0323 (0.007)	0.058 (0.0065)	0.0485 (0.0072)	0.0616 (0.0076)	0.0496 (0.0074)	0.0551 (0.007)	0.0462 (0.0068)
BK	0.0906 (0.0023)	0.211 (0.011)	0.164 (0.0077)	0.138 (0.0087)	0.139 (0.0081)	0.082 (0.0086)	0.0647 (0.0084)	0.0511 (0.0076)	0.0833 (0.0083)	0.077 (0.0074)	0.0613 (0.0078)	0.0647 (0.0077)	0.06 (0.0079)	0.0563 (0.0074)	0.0478 (0.0078)
C	0.0611 (0.002)	0.0965 (0.0091)	0.0592 (0.0077)	0.0764 (0.0081)	0.0696 (0.0075)	0.0451 (0.008)	0.0248 (0.0065)	0.0399 (0.0066)	0.0535 (0.0067)	0.0424 (0.0067)	0.0493 (0.0067)	0.0572 (0.0065)	0.0556 (0.0074)	0.102 (0.0074)	0.0924 (0.0074)
CMCSA	0.152 (0.0032)	0.389 (0.015)	0.315 (0.011)	0.197 (0.0089)	0.263 (0.011)	0.37 (0.012)	0.14 (0.0084)	0.0852 (0.0089)	0.0742 (0.0077)	0.0683 (0.0076)	0.082 (0.0082)	0.0589 (0.0072)	0.0356 (0.0069)	0.0694 (0.0077)	0.0521 (0.008)
COF	0.0953 (0.0026)	0.346 (0.014)	0.193 (0.011)	0.134 (0.0086)	0.153 (0.0088)	0.1 (0.0084)	0.0595 (0.0074)	0.0366 (0.009)	0.0636 (0.0082)	0.055 (0.0081)	0.0517 (0.0074)	0.0784 (0.0089)	0.0636 (0.0079)	0.0371 (0.007)	0.0342 (0.0072)
CPB	0.113 (0.0024)	0.212 (0.011)	0.177 (0.0079)	0.172 (0.0088)	0.201 (0.0085)	0.127 (0.0082)	0.0934 (0.0081)	0.0893 (0.0078)	0.0907 (0.0085)	0.0798 (0.008)	0.0672 (0.0084)	0.0713 (0.0082)	0.0774 (0.0077)	0.0765 (0.0082)	0.0784 (0.0082)
CSCO	0.0576 (0.002)	0.0912 (0.0088)	0.0696 (0.0075)	0.0635 (0.0073)	0.0427 (0.0069)	0.0378 (0.0066)	0.0525 (0.0075)	0.0644 (0.0072)	0.0585 (0.0071)	0.0684 (0.007)	0.0725 (0.0071)	0.0488 (0.0076)	0.0418 (0.0073)	0.0554 (0.0074)	0.0489 (0.0068)
DELL	0.0663 (0.0021)	0.0603 (0.0088)	0.0401 (0.0078)	0.101 (0.0079)	0.0997 (0.0076)	0.0483 (0.0068)	0.0501 (0.0082)	0.0653 (0.0076)	0.0638 (0.0067)	0.0616 (0.0076)	0.0791 (0.0084)	0.0677 (0.0071)	0.0654 (0.0068)	0.0553 (0.0086)	0.0677 (0.0081)
GOOG	0.0593 (0.0031)								0.0709 (0.012)	0.0644 (0.008)	0.0663 (0.0074)	0.0653 (0.008)	0.0559 (0.0075)	0.0533 (0.0077)	0.0464 (0.0079)
GS	0.0598 (0.0023)			0.167 (0.01)	0.106 (0.0081)	0.068 (0.0081)	0.0498 (0.0074)	0.0472 (0.0075)	0.0523 (0.0069)	0.0401 (0.0073)	0.031 (0.0071)	0.0534 (0.0079)	0.0399 (0.0068)	0.0448 (0.0066)	0.0548 (0.0079)
HNZ	0.114 (0.0025)	0.216 (0.011)	0.207 (0.0088)	0.191 (0.0091)	0.2 (0.0094)	0.131 (0.0091)	0.0754 (0.0082)	0.0738 (0.008)	0.0738 (0.0077)	0.0925 (0.0077)	0.0864 (0.0089)	0.0695 (0.0081)	0.0511 (0.0076)	0.0798 (0.0078)	0.0731 (0.0088)
HPQ	0.0727 (0.0021)	0.137 (0.0097)	0.104 (0.0083)	0.116 (0.0078)	0.0701 (0.0076)	0.0775 (0.0083)	0.0782 (0.0081)	0.0715 (0.008)	0.0538 (0.0069)	0.0526 (0.0074)	0.0602 (0.008)	0.066 (0.0072)	0.0574 (0.0071)	0.0531 (0.0079)	0.0392 (0.0078)
IBM	0.0511 (0.002)	0.0877 (0.0095)	0.0902 (0.0072)	0.0598 (0.0072)	0.0649 (0.0078)	0.0369 (0.0074)	0.0323 (0.0069)	0.0342 (0.0059)	0.0397 (0.0072)	0.0432 (0.0071)	0.0444 (0.007)	0.059 (0.0075)	0.0323 (0.0076)	0.0548 (0.0075)	0.0453 (0.007)
INTC	0.0519 (0.0019)	0.061 (0.0089)	0.0676 (0.0075)	0.0502 (0.0068)	0.0552 (0.0069)	0.0427 (0.0068)	0.0402 (0.0069)	0.049 (0.0071)	0.0489 (0.0062)	0.0554 (0.0068)	0.0661 (0.0068)	0.0553 (0.007)	0.0492 (0.0067)	0.0492 (0.0071)	0.0385 (0.0071)
JPM	0.0648 (0.0021)	0.145 (0.0092)	0.0926 (0.0072)	0.0807 (0.0079)	0.0788 (0.0076)	0.0741 (0.0077)	0.047 (0.0075)	0.049 (0.0073)	0.0429 (0.0066)	0.0487 (0.0075)	0.0548 (0.0078)	0.0589 (0.0079)	0.0466 (0.0067)	0.0537 (0.007)	0.0553 (0.008)
KFT	0.115 (0.0032)					0.172 (0.014)	0.132 (0.0083)	0.123 (0.0085)	0.126 (0.0089)	0.141 (0.0094)	0.115 (0.01)	0.0914 (0.0088)	0.064 (0.0073)	0.0568 (0.013)	
KO	0.0765 (0.0021)	0.144 (0.01)	0.109 (0.0076)	0.108 (0.0076)	0.109 (0.0067)	0.0938 (0.0076)	0.0505 (0.0074)	0.0543 (0.0074)	0.0673 (0.0069)	0.0755 (0.0075)	0.0543 (0.0071)	0.0645 (0.008)	0.0577 (0.0076)	0.06 (0.0075)	0.0421 (0.0075)
MON	0.127 (0.004)				0.456 (0.027)	0.439 (0.014)	0.245 (0.014)	0.156 (0.0096)	0.0805 (0.0086)	0.0557 (0.0086)	0.0427 (0.0079)	0.0542 (0.0081)	0.0419 (0.0072)	0.0532 (0.0077)	0.0547 (0.0076)
MSFT	0.0583 (0.002)	0.0662 (0.0081)	0.0803 (0.0082)	0.0783 (0.0072)	0.0537 (0.0074)	0.0449 (0.0073)	0.0518 (0.0066)	0.0428 (0.0069)	0.0512 (0.0067)	0.0652 (0.0073)	0.0722 (0.0075)	0.0549 (0.0075)	0.0375 (0.0073)	0.0648 (0.0073)	0.000146 (0.027)
T	0.0833 (0.0022)	0.176 (0.011)	0.13 (0.008)	0.0994 (0.0083)	0.117 (0.0083)	0.0815 (0.0078)	0.0712 (0.0074)	0.0768 (0.008)	0.08 (0.008)	0.0813 (0.0079)	0.0526 (0.0079)	0.0677 (0.0081)	0.052 (0.0073)	0.0446 (0.0074)	0.0615 (0.0072)
XOM	0.0452 (0.0021)			0.0583 (0.018)	0.0747 (0.0079)	0.0569 (0.008)	0.0256 (0.007)	0.0499 (0.0067)	0.0576 (0.0074)	0.0445 (0.0065)	0.0225 (0.007)	0.0326 (0.0065)	0.0373 (0.0068)	0.0609 (0.0068)	0.0335 (0.0072)
Finance	0.0725 (0.016)	0.183 (0.1)	0.119 (0.057)	0.114 (0.037)	0.111 (0.033)	0.0729 (0.018)	0.0464 (0.015)	0.0432 (0.0068)	0.0546 (0.018)	0.0535 (0.013)	0.0494 (0.01)	0.0624 (0.0087)	0.0526 (0.0088)	0.0582 (0.023)	0.0551 (0.02)
Food	0.109 (0.019)	0.19 (0.041)	0.164 (0.05)	0.157 (0.044)	0.241 (0.15)	0.193 (0.14)	0.119 (0.076)	0.0993 (0.04)	0.0878 (0.023)	0.089 (0.032)	0.073 (0.028)	0.0702 (0.014)	0.0584 (0.013)	0.0653 (0.012)	0.0621 (0.017)
Tech	0.0729 (0.03)	0.134 (0.1)	0.115 (0.081)	0.0985 (0.045)	0.0941 (0.067)	0.0893 (0.11)	0.0661 (0.032)	0.0646 (0.019)	0.0604 (0.012)	0.0621 (0.01)	0.0638 (0.014)	0.0588 (0.008)	0.0465 (0.011)	0.0562 (0.0074)	0.0463 (0.019)
All	0.0797 (0.028)	0.158 (0.094)	0.125 (0.069)	0.11 (0.046)	0.128 (0.096)	0.108 (0.11)	0.0712 (0.05)	0.066 (0.031)	0.0649 (0.02)	0.0651 (0.022)	0.0601 (0.02)	0.0612 (0.012)	0.0505 (0.012)	0.059 (0.014)	0.0512 (0.019)

42

Stock	$C_{BV}^{(d)}$ (Daily)			$C_{BV}^{(w)}$ (Weekly)			$C_{BV}^{(m)}$ (Monthly)		
	Tech	Fin	Food	Tech	Fin	Food	Tech	Fin	Food
AAPL	0.0679 (0.011)	0.0538 (0.012)	0.0678 (0.018)	0.137 (0.019)	0.116 (0.019)	0.131 (0.026)	0.211 (0.043)	0.224 (0.045)	0.28 (0.028)
CSCO	0.0595 (0.0082)	0.0381 (0.0044)	0.011 (0.016)	0.105 (0.021)	0.0563 (0.016)	0.00573 (0.019)	0.143 (0.032)	0.0459 (0.041)	−0.0357 (0.061)
DELL	0.0461 (0.0085)	0.0405 (0.0074)	0.0275 (0.013)	0.0876 (0.023)	0.117 (0.015)	0.038 (0.016)	0.0766 (0.037)	0.153 (0.048)	0.00765 (0.046)
GOOG	0.0331 (0.0098)	0.0372 (0.019)	0.0182 (0.011)	0.0915 (0.025)	0.0647 (0.031)	0.0924 (0.039)	0.0992 (0.043)	0.0981 (0.065)	0.153 (0.0093)
HPQ	0.0571 (0.012)	0.055 (0.0085)	0.0765 (0.015)	0.105 (0.03)	0.12 (0.023)	0.185 (0.027)	0.172 (0.051)	0.207 (0.051)	0.347 (0.039)
IBM	0.0666 (0.0039)	0.0613 (0.014)	0.064 (0.016)	0.105 (0.013)	0.104 (0.021)	0.0913 (0.035)	0.112 (0.03)	0.226 (0.047)	0.121 (0.075)
INTC	0.0519 (0.01)	0.0397 (0.011)	0.0214 (0.0092)	0.09 (0.023)	0.0643 (0.012)	0.0339 (0.0093)	0.107 (0.038)	0.0847 (0.031)	0.0441 (0.032)
MSFT	0.0521 (0.0094)	0.0449 (0.0095)	0.018 (0.017)	0.109 (0.012)	0.0874 (0.016)	0.0504 (0.032)	0.163 (0.014)	0.132 (0.0065)	0.0957 (0.055)
T	0.0495 (0.02)	0.0618 (0.016)	0.0889 (0.017)	0.101 (0.029)	0.129 (0.042)	0.17 (0.047)	0.155 (0.058)	0.257 (0.052)	0.35 (0.075)
BAC	0.0639 (0.0077)	0.101 (0.019)	0.0628 (0.015)	0.12 (0.015)	0.217 (0.019)	0.148 (0.046)	0.241 (0.033)	0.342 (0.015)	0.304 (0.079)
BK	0.0776 (0.015)	0.11 (0.019)	0.0984 (0.02)	0.16 (0.033)	0.212 (0.031)	0.208 (0.042)	0.225 (0.041)	0.284 (0.045)	0.255 (0.061)
C	0.0261 (0.0082)	0.069 (0.0078)	0.0163 (0.021)	0.0553 (0.016)	0.13 (0.019)	0.0059 (0.05)	0.0958 (0.043)	0.268 (0.031)	0.0859 (0.073)
COF	0.0763 (0.023)	0.133 (0.026)	0.123 (0.028)	0.181 (0.044)	0.244 (0.053)	0.235 (0.061)	0.276 (0.061)	0.415 (0.049)	0.369 (0.1)
GS	0.0424 (0.0096)	0.0901 (0.01)	0.0543 (0.014)	0.0828 (0.021)	0.167 (0.033)	0.108 (0.022)	0.145 (0.038)	0.246 (0.045)	0.23 (0.032)
JPM	0.063 (0.0078)	0.125 (0.015)	0.0743 (0.015)	0.127 (0.022)	0.187 (0.024)	0.135 (0.027)	0.195 (0.045)	0.301 (0.043)	0.32 (0.037)
CPB	0.0685 (0.02)	0.0989 (0.0085)	0.135 (0.036)	0.129 (0.033)	0.199 (0.028)	0.236 (0.071)	0.258 (0.05)	0.35 (0.021)	0.373 (0.096)
HNZ	0.0839 (0.021)	0.0827 (0.023)	0.137 (0.036)	0.145 (0.048)	0.188 (0.047)	0.237 (0.062)	0.231 (0.072)	0.341 (0.052)	0.472 (0.046)
KFT	0.0253 (0.011)	0.0172 (0.015)	0.0721 (0.014)	0.0645 (0.027)	0.0237 (0.038)	0.105 (0.04)	0.0971 (0.067)	0.0874 (0.05)	0.231 (0.08)
KO	0.0667 (0.013)	0.0706 (0.0087)	0.0861 (0.02)	0.13 (0.025)	0.122 (0.019)	0.153 (0.05)	0.244 (0.051)	0.268 (0.054)	0.312 (0.11)
MON	0.0648 (0.042)	0.0369 (0.019)	0.1 (0.0088)	0.13 (0.066)	0.0713 (0.051)	0.191 (0.029)	0.156 (0.095)	0.149 (0.064)	0.333 (0.046)
XOM	0.0492 (0.0078)	0.0499 (0.012)	0.0573 (0.0069)	0.0976 (0.013)	0.0935 (0.018)	0.088 (0.0072)	0.0544 (0.034)	0.0934 (0.028)	0.134 (0.021)

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 $C_{BV}^{(d)}$ 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 apparently with food and finance stocks.

Notation	Description	Estimation Method
$\sigma^{(i)}$	Daily diffusive volatility for stock i. Constant across time	Bipower variation
ρ_d	Diffusive correlation between two stocks.	Bipower correlation (Equation (20))
ρ_c	Correlation in cojump size	Sample correlation after identifying cojumps using Equation (21)
$\sigma_{\text{jump}}^{(i)}$	Standard deviation of idiosyncratic jump size for stock i	Sample standard deviation after identifying jumps
$\sigma_{cj}^{(i)}$	Standard deviation of cojump size for stock i	Sample standard deviation after identifying cojumps
$\lambda^{(i)}$	Frequency parameter of idiosyncratic jumps for stock i in day^{-1}	Jiang-Oomen test at 0.5 percent significance level
λ_c	Frequency parameter of idiosyncratic jumps for stock i in day^{-1}	Jiang-Oomen test at 0.5 percent significance level

Table 4: Notation summary and estimation methods

Parameter	Value
$\sigma^{(i)}$	0.01
$\lambda^{(i)}$	0.25
$\sigma_c \equiv \sigma_{cj}^{(i)} = \sigma_{\text{jump}}^{(i)}$	0.025
ρ_d	$\{0, 0.20, 0.40, 0.60, 0.80\}$
ρ_c	$\{0, 0.25, 0.50, 0.75\}$
λ_c	$\{0.075, 0.25, 1\}$

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

	Daily	Weekly	Monthly
C_{BV}^{sim}	0.0553 (0.037)	0.0815 (0.091)	0.0857 (0.16)
C_{BV}^{data}	0.0658	0.18	0.305
C_{MinV}^{sim}	0.0451 (0.031)	0.07 (0.078)	0.0775 (0.14)
C_{MinV}^{data}	0.0793	0.153	0.222
C_{MedV}^{sim}	0.0553 (0.036)	0.0799 (0.088)	0.0869 (0.15)
C_{MedV}^{data}	0.0825	0.173	0.249

(a) BAC and C

	Daily	Weekly	Monthly
C_{BV}^{sim}	0.0554 (0.038)	0.105 (0.087)	0.138 (0.14)
C_{BV}^{data}	0.0605	0.137	0.144
C_{MinV}^{sim}	0.0373 (0.032)	0.0876 (0.08)	0.126 (0.14)
C_{MinV}^{data}	0.0469	0.112	0.0582
C_{MedV}^{sim}	0.0532 (0.037)	0.105 (0.085)	0.141 (0.14)
C_{MedV}^{data}	0.0663	0.12	0.0972

(b) BK and KFT

Table 6: Comparison of simulated and observed values of $C_{\times}$ for (a) BAC and C and (b) 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 $C_{\times}$ calculated in the simulation, using the 1000 sample paths. In (a), there is a noticable 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. In (b), unlike in (a), there is no systematic discrepancy between the simulations and the data.

Industry	Quantity	Daily	Weekly	Monthly
Fin/Fin	C_{BV}^{sim}	0.0614 (0.018)	0.114 (0.056)	0.138 (0.078)
	C_{BV}^{data}	0.103 (0.031)	0.187 (0.061)	0.294 (0.076)
Food/Food	C_{BV}^{sim}	0.0295 (0.014)	0.0567 (0.042)	0.0663 (0.058)
	C_{BV}^{data}	0.102 (0.051)	0.189 (0.11)	0.349 (0.17)
Tech/Tech	C_{BV}^{sim}	0.0709 (0.036)	0.163 (0.12)	0.217 (0.17)
	C_{BV}^{data}	0.05 (0.021)	0.0995 (0.056)	0.133 (0.096)

(a) Within-industry comparisons

Industry	Quantity	Daily	Weekly	Monthly
Fin/Food	C_{BV}^{sim}	0.0266 (0.022)	0.0562 (0.062)	0.0721 (0.083)
	C_{BV}^{data}	0.0587 (0.045)	0.126 (0.11)	0.239 (0.15)
Fin/Tech	C_{BV}^{sim}	0.0376 (0.025)	0.0769 (0.079)	0.095 (0.11)
	C_{BV}^{data}	0.0434 (0.024)	0.0912 (0.051)	0.146 (0.11)
Food/Tech	C_{BV}^{sim}	0.0372 (0.04)	0.0935 (0.12)	0.123 (0.17)
	C_{BV}^{data}	0.0353 (0.038)	0.0757 (0.082)	0.128 (0.16)

(b) Out-of-industry comparisons

Table 7: Average values for C_{BV}^{sim} and C_{BV}^{data} at daily, weekly, and monthly levels, grouping by industry pairs. Note that "Fin" refers to the finance industry. The numbers in parentheses are sample standard deviations of all the pairs in a particular group.

Industry Comparison	Daily	Weekly	Monthly
Finance/Finance	0.0451 (0.01)	0.11 (0.021)	0.168 (0.034)
Food/Food	0.0728 (0.017)	0.146 (0.033)	0.301 (0.045)
Tech/Tech	0.0384 (0.0057)	0.113 (0.018)	0.18 (0.026)
Finance/Food	0.0119 (0.0039)	0.0341 (0.012)	0.0605 (0.023)
Finance/Tech	0.00437 (0.0015)	0.0156 (0.005)	0.0247 (0.0087)
Food/Tech	0.0178 (0.0054)	0.045 (0.017)	0.0845 (0.02)

Table 8: Average values for $\left| C_{BV}^{\text{sim}} - C_{BV}^{\text{data}} \right|$ 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 ρ and τ, show similar patterns.

Industry	Daily	Weekly	Monthly	
Finance/Finance	2.09	2.19	1.74	2.01
Food/Food	7.37	6.15	5.64	6.36
Tech/Tech	0.92	1.17	1.15	1.08
	2.51	2.42	2.21	2.38

(a) Within-industry comparisons

Industry	Daily	Weekly	Monthly	
Finance/Food	4.13	4.36	3.31	4.10
Finance/Tech	0.62	0.65	0.87	0.72
Food/Tech	1.07	1.63	1.59	1.43
	1.62	1.91	1.73	1.75

(b) Out-of-industry comparisons

Table 9: Average values of the χ^2 statistic, computed for each pair of stocks and averaged over all pairs in the industry and time grouping. The final column in each table corresponds to grouping across the entire industry pair, regardless of the time interval used; the last row corresponds to grouping across the entire time interval, regardless of the industry pair. Note that these values should be not compared against a χ_1^2 distribution to test for significance, as discussed in Section 7.2.

D Figures
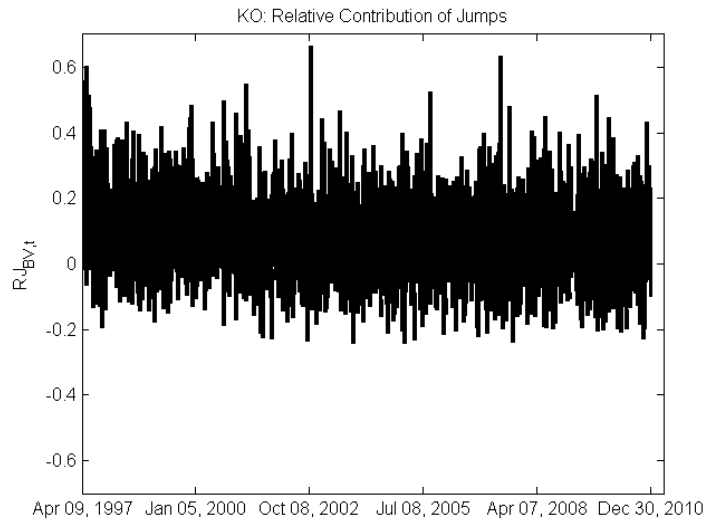
Figure 1: RJ_{BV} time series for a single stock (KO). The statistic is calculated at the daily level using 5-minute intervals and is plotted for days on which we have price data for KO. There is a considerable amount of variability in this process, and a slight time trend may be present.

(a) BAC

(b) DELL
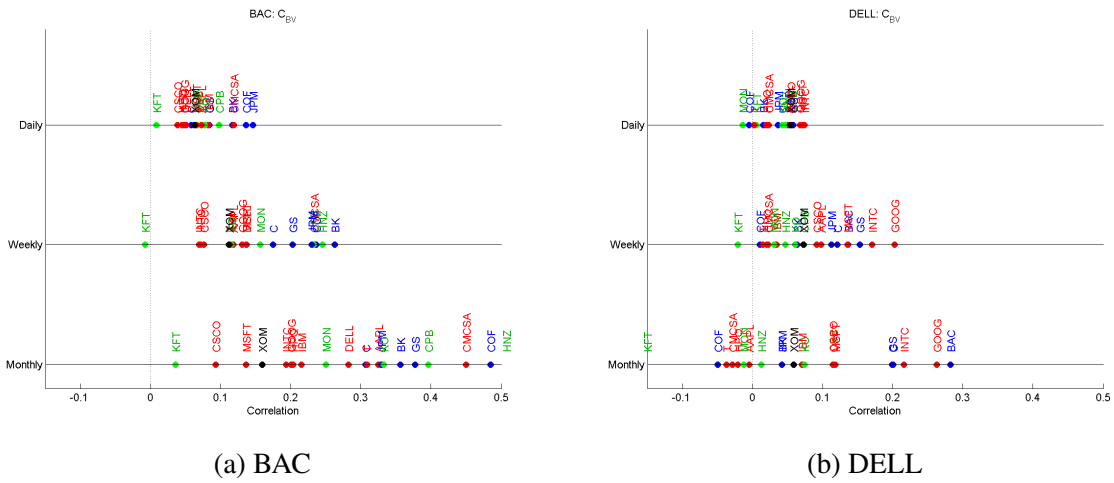
Figure 2: Values of $C_{BV}^{(d)}$, $C_{BV}^{(w)}$, and $C_{BV}^{(m)}$ for BAC and DELL. The points in red represent tech stocks, those in blue are finance stocks, and those in green are food/agricultural stocks. The general observations we find—more clear in the figure for BAC—is that pairs of stocks in the same industry tend to have higher values of C_{BV}. Furthermore, the magnitude of C_{BV} increases with the time frame considered.
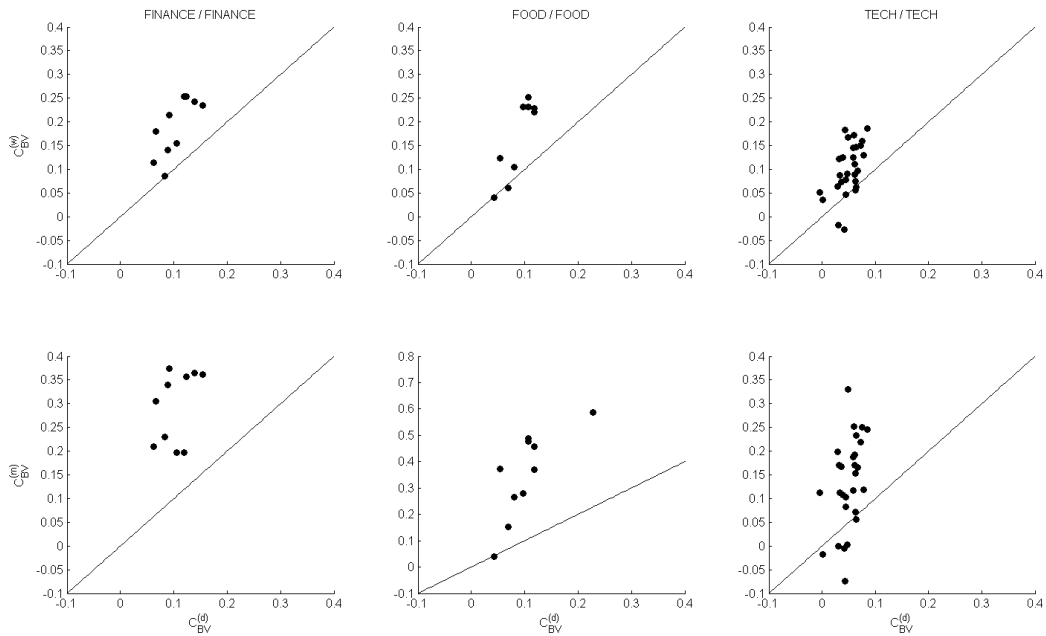
(a) Pairs within the same industry
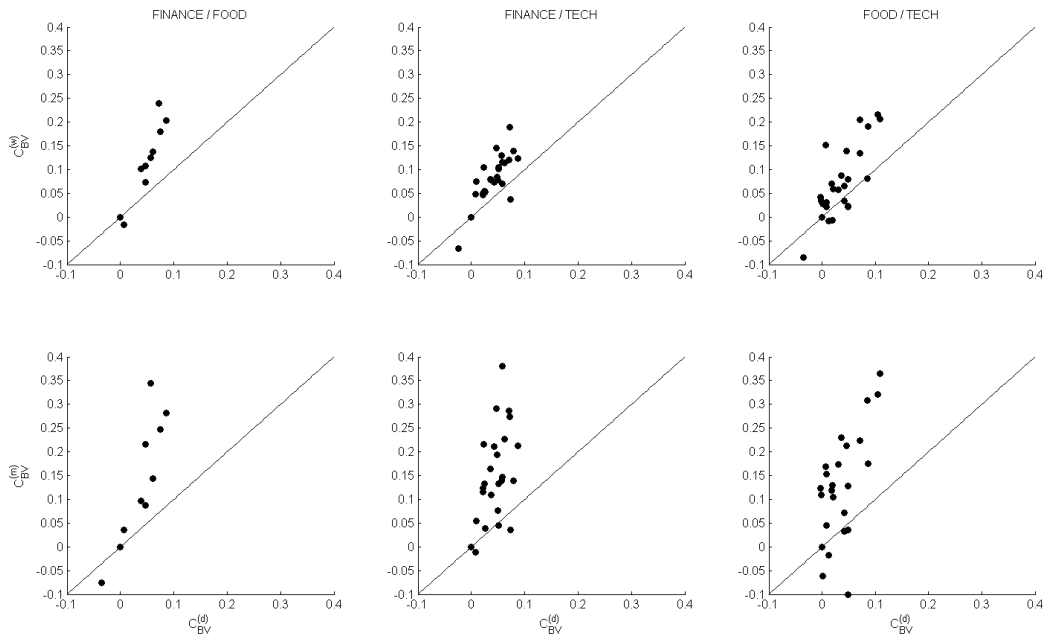
(b) Pairs from different industries

Figure 3: Comparisons of $C_{BV}^{(w)}$ vs. $C_{BV}^{(d)}$ and $C_{BV}^{(m)}$ vs. $C_{BV}^{(d)}$ when considering (a) pairs of stocks from the same industry or (b) pairs of stocks from different industries. The first row in each group plots $C_{BV}^{(w)}$ vs. $C_{BV}^{(d)}$ and the second plots $C_{BV}^{(m)}$ vs. $C_{BV}^{(d)}$. Scales are the same in each plot, except for the FOOD/FOOD plot when comparing $C_{BV}^{(m)}$ to $C_{BV}^{(d)}$. The solid black line is a $45°$ line. Each dot in the scatter plot represents a particular pair of stocks.
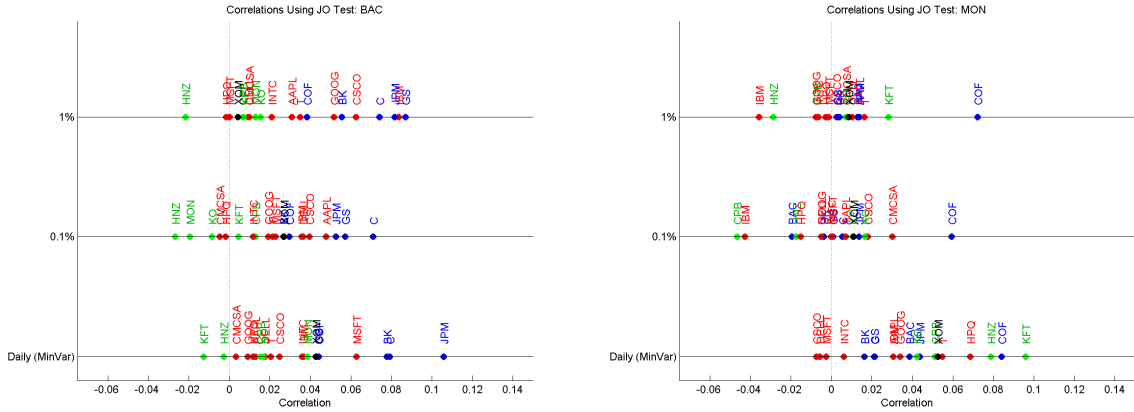
(a) JC at 1 and 0.1 percent levels vs. $C_{MinV}^{(d)}$ for BAC

(b) JC at 1 and 0.1 percent levels vs. $C_{MinV}^{(d)}$ for MON
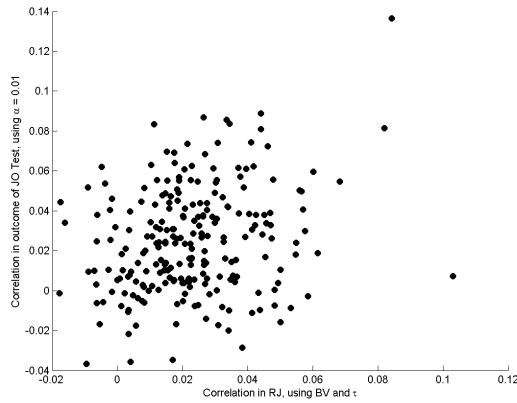
(c) JC at 1 percent level vs. $\tau_{BV}^{(d)}$

Figure 4: Comparisons of JC at 1 and 0.1 percent levels against variations of $C_{\times}^{(d)}$. 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 $C_{MinV}^{(d)}$. Panel (c) plots JC against $\tau_{BV}^{(d)}$ for all pairs of stocks considered. We see no clear positive association between these statistics.
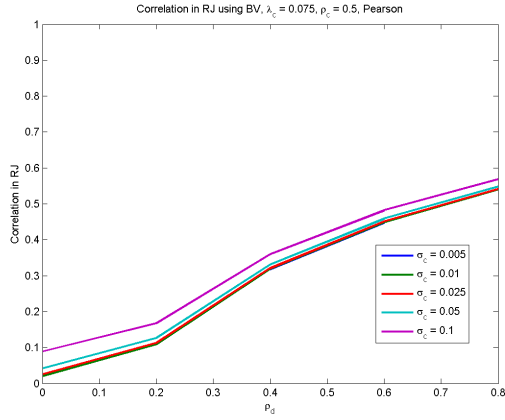
(a) $\lambda_c = 0.075$
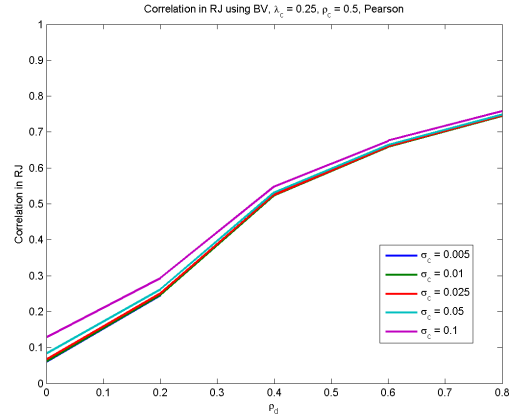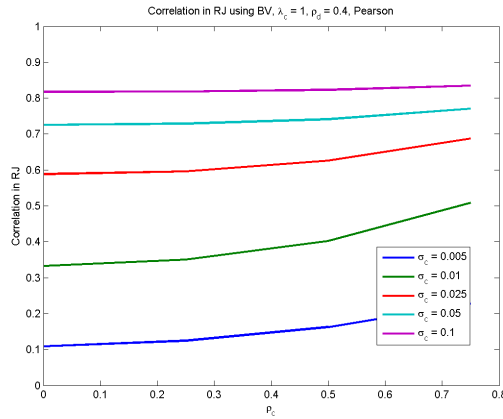
(b) $\lambda_c = 0.25$
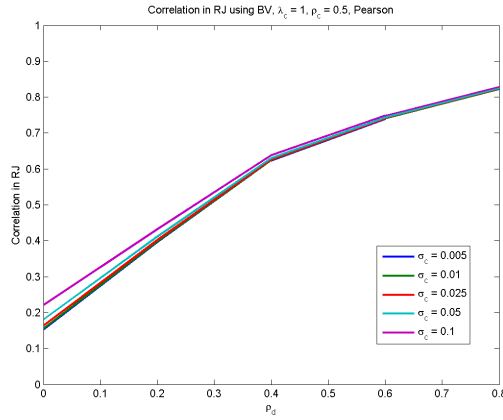
(c) $\lambda_c = 1$

Figure 5: C_{BV} as a function of ρ_d for various values of σ_c and λ_c, fixing $\rho_c = 0.5$. The different panels give different values of λ_c whereas the different colors within a panel correspond to different σ_c. Standard errors on the mean values calculated through simulation are small and thus ignored in the plots. The general observation is that ρ_d has a strong effect on C_{BV}.
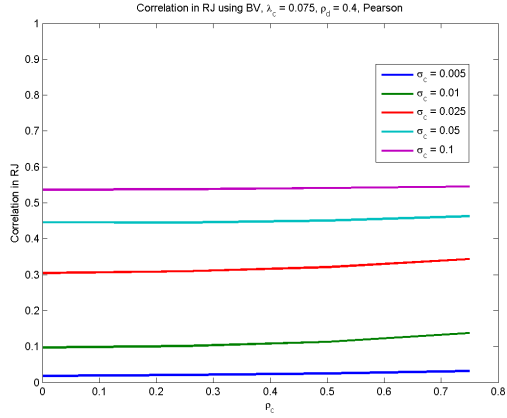
(a) $\lambda_c = 0.075$
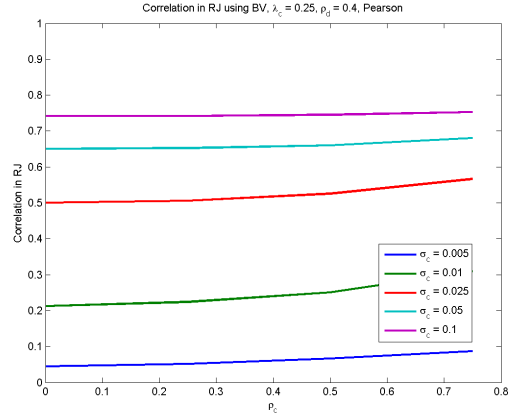
(b) $\lambda_c = 0.25$

(c) $\lambda_c = 1$

Figure 6: C_{BV} as a function of ρ_c for various values of σ_c and λ_c, fixing $\rho_d = 0.4$. The different panels give different values of λ_c whereas the different colors within a panel correspond to different σ_c. Standard errors on the mean values calculated through simulation are small and thus ignored in the plots. The general observation is that ρ_c has a very small effect on C_{BV}, although for sufficiently small σ_c, the effect is noticeable at high ρ_c.
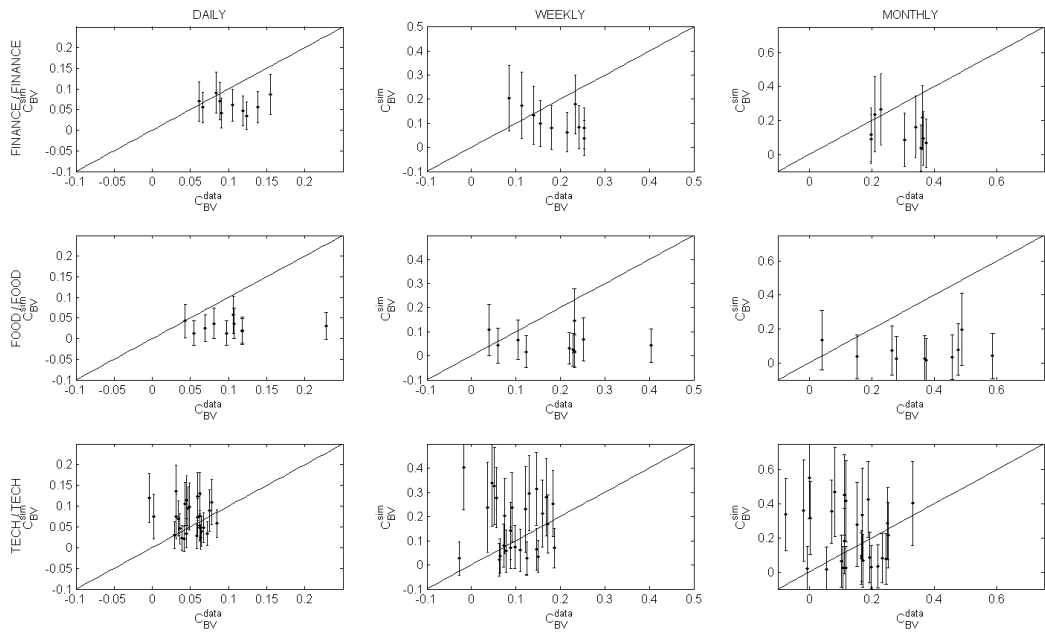
(a) Pairs within the same industry
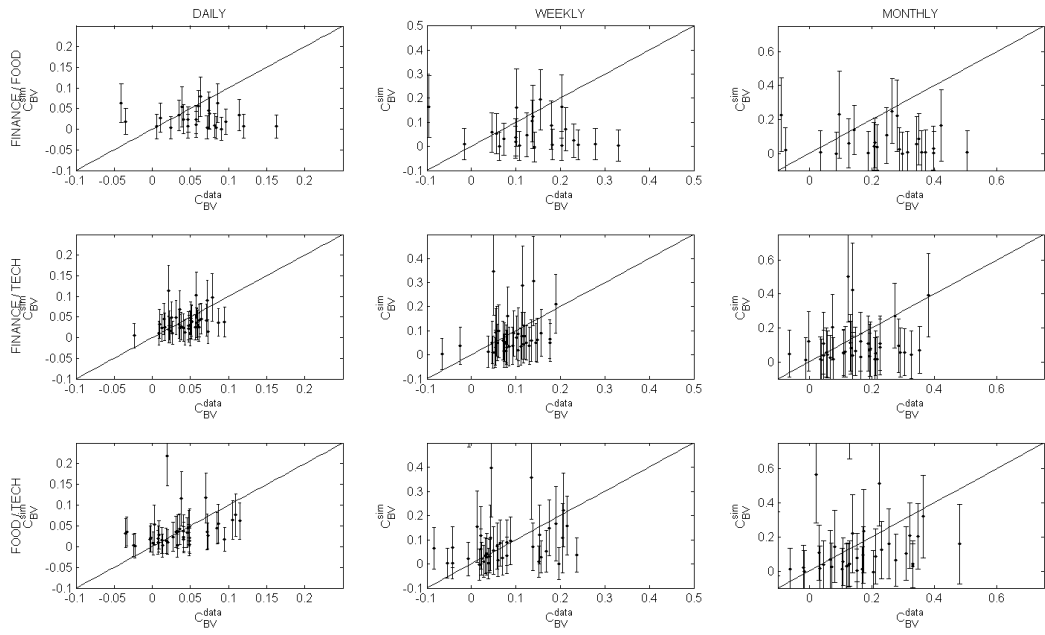
(b) Pairs from different industries

Figure 7: Comparisons of C_{BV}^{data} vs. C_{BV}^{sim} at the daily, weekly, and monthly levels. The first column shows the comparisons at the daily level, the second at the weekly level, and the third at the monthly level. Each row corresponds to a particular industry pair. The horizontal axis on each plot is the value estimated from the data, and the vertical axis is that obtained from simulation. The error bars indicate one standard deviation in the simulated runs.
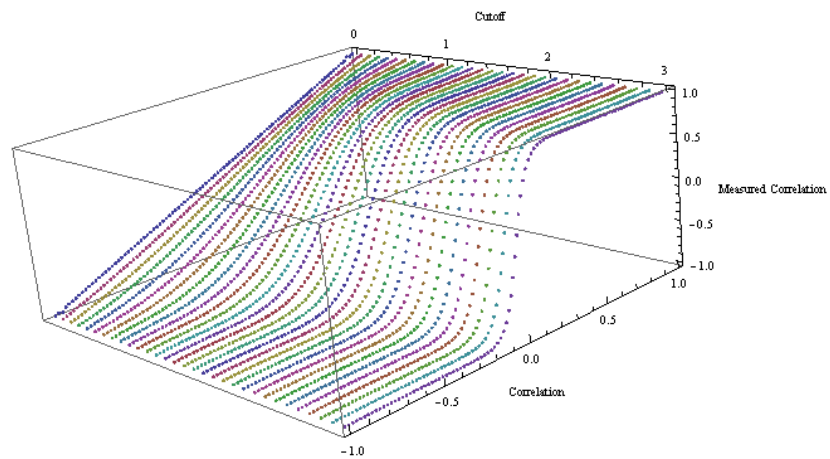
Figure 8: Measured correlation of two jointly normal variables with correlation ρ (plotted on the "Correlation" axis) when observing only the realizations that are above some cutoff c (plotted on the "Cutoff" axis). The value of the function is calculated through numerical integration of Equation (22).

References

Ait-Sahalia, Y., & Jacod, J. (2009). Testing for Jumps in a Discretely Observed Process. *Annals of Statistics*, *37*(1), 184–222.

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

Andersen, T., Bollerslev, T., & Diebold, F. (2007). Roughing it Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Review of Economics and Statistics*, *84*(4), 701–720.

Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2000). Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian. *Multinational Finance Journal*, *4*(3), 159–179.

Andersen, T., Bollerslev, T., & Dobrev, D. (2006). No-Arbitrage and Semi-Martingale Restrictions for Continuous-Time Volatility Models Subject to Leverage Effects, Jumps, and iid Noise: Theory and Testable Distributional Implications.
(NBER Working Paper)

Andersen, T., Dobrev, D., & Schaumburg, E. (2010). Jump-Robust Volatility Estimation using Nearest Neighbor Truncation.
(Federal Reserve Bank of New York Staff Report 465)

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

Barndorff-Nielsen, O., & Shephard, N. (2005). Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. *Journal of Financial Econometrics*, *4*(1), 1–30.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Bollerslev, T., Kretschmer, U., Pigorsch, C., & Tauchen, G. (2009). A Discrete-Time Model for Daily SP500 Returns and Realized Variations: Jumps and Leverage Effects. *Journal of Econometrics*, *150*(2), 151–166.

Bollerslev, T., Todorov, V., & Li, S. Z. (2011). Jump Tails, Extreme Dependencies, and the

Distribution of Stock Returns. *Journal of Econometrics*. (Forthcoming)

Chan, W. H., & Maheu, J. M. (2002). Conditional Jump Dynamics in Stock Market Returns. *Journal of Business and Economic Statistics*, *20*(2), 377–389.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

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

Huang, X., & Tauchen, G. (2005). The Relative Contribution of Jumps to Total Price Variance. *Journal of Financial Econometrics*, *3*(4), 256–499.

Jiang, G., & Oomen, R. (2008). Testing for Jumps When Asset Prices are Observed with Noise: A Swap Variance Approach. *Journal of Econometrics*, *144*(2), 352–370.

Kuo, S. (2002). A Jump-Diffusion Model for Option Pricing. *Management Science*, *48*(8), 1086–1101.

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

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, *7*, 77–91.

Merton, R. (1976). Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Econometrics*, *3*(1), 125–144.

Rognlie, M. (2010). Spurious Jump Detection and Intraday Changes in Volatility. (Duke University Economics Honors Thesis)

Royen, T. (1990). Expansions for the Multivariate Chi-Square Distribution. *Journal of Multivariate Analysis*, *38*(2), 213-232.