High-Frequency Returns, Jumps and the Mixture of Normals Hypothesis

Jeff Fleming and Bradley S. Paye
Jones Graduate School of Management, Rice University

November 18, 2005

ABSTRACT

Previous empirical studies find both evidence of jumps in asset prices and that returns standardized by 'realized volatility' are approximately standard normal. These findings appear to be contradictory. This paper resolves the apparent contradiction using high-frequency returns data for 20 heavily-traded US stocks. We show that microstructure noise may bias kurtosis estimates for standardized returns. When we apply a bias-corrected realized variance estimator we find that standardized returns are platykurtotic and that the standard normal distribution is easily rejected. We show that when daily returns are standardized using realized bipower variation, an estimator for integrated volatility that is robust to the presence of jumps, the resulting series does appear to be approximately standard normal. For many of the stocks we consider the standardized returns are statistically indistinguishable from the standard normal distribution when confronted with a set of moment-based hypothesis tests. Our results suggest that there is no empirical contradiction: jumps should be included in stock price models.

We thank Scott Baggett for helping us construct the price datasets from TAQ data and Barbara Ostdiek for suggesting useful references. We also received helpful comments from members of the Finance Group at Rice University and seminar participants at the Rady School of Management at the University of California, San Diego. Finally, we thank James Weston for providing useful insights and advice regarding TAQ data.

1. Introduction

Asset prices are frequently modeled as evolving according to a continuous diffusive process. This class of models is both tractable and flexible. Indeed, by superimposing multiple stochastic volatility diffusive processes, such models may accommodate many of the observed features of asset returns, such as time-variation in both conditional volatility and conditional kurtosis.[1] Nevertheless, the assumption that prices exhibit continuous sample paths is restrictive. It is certainly plausible that asset prices exhibit occasional jumps, possibly related to the release of firm-specific or macroeconomic news. Such reasoning suggests that asset prices may be more appropriately modeled as jump diffusions. While jump diffusion models permit greater flexibility than continuous diffusive processes, there are potential drawbacks to permitting jump behavior. Jump diffusions are more difficult to estimate. Further, jumps may pose different risks for investors and such risks may be difficult to price empirically using cross-sectional data.

Given the additional difficulties of incorporating jumps into asset price models, assessing the empirical evidence regarding both the existence and importance of jumps in asset prices is a crucial task. A number of recent papers address this question. Andersen, Benzoni and Lund (2001) and Chernov, Gallant, Ghysels and Tauchen (2003) find evidence of jumps based on parametric estimation of jump diffusion models. More recently, Andersen, Bollerslev and Diebold (2004), Huang and Tauchen (2005), Tauchen and Zhou (2005) and Barndorff-Nielsen and Shephard (2006, 2005) provide evidence of jumps using techniques that exploit the information in high-frequency intraday asset returns.

On the other hand, a number of recent studies observe that daily returns are nearly Gaussian when standardized by "realized volatility," constructed using high-frequency intraday returns.[2] As shown by Andersen, Bollerslev, Diebold and Labys (2003), this finding is consistent with a simple continuous diffusion model for asset prices. The finding that daily returns standardized by realized volatility are Gaussian appears starkly at odds with the emerging body of literature that directly tests for jumps in asset prices.

This paper resolves the apparent empirical contradiction, at least for the case of returns on 20 heavily-traded US stocks. We first show that daily returns for these stocks, standardized by realized volatility, are in fact *not* Gaussian. In particular, the standardized returns appear to be platykurtotic. We demonstrate that kurtosis statistics for standardized

[1] See, e.g., Chernov, Gallant, Ghysels and Tauchen (2003).

[2] For example, Andersen, Bollerslev, Diebold and Ebens (2001) find that stock returns standardized by realized variance are approximately Gaussian and Andersen, Bollerslev, Diebold and Labys (2001) obtain a similar finding for exchange rate returns.

returns constructed using the standard realized variance estimator may be biased and misleading. When microstructure effects induce negative serial correlation in high-frequency returns, the corresponding upward bias in realized variance induces an associated upward bias in the kurtosis estimate for standardized returns. We illustrate that the upward bias in kurtosis estimates is due to the fact that microstructure effects tend to be weaker on days that exhibit large returns. While the realized volatility is biased upward for all returns, it is less biased upward for extreme returns, leading to an upward bias in the estimated kurtosis for standardized returns.

When we base kurtosis estimates on a bias-corrected version of the realized variance, we find that the normality assumption is resoundingly rejected for nearly all of the stocks. Further, this result is robust to sampling frequencies from one minute to 30 minutes. While we concur with Andersen, Bollerslev, Diebold and Ebens that standardized returns are much closer to normally distributed than raw daily returns, we statistically reject that standardized returns are in fact normal.

A natural explanation for the non-normality of returns standardized by realized variance is the presence of jumps. We explore whether accounting for jumps leads to standardized returns that do appear to be Gaussian. We apply the realized bipower variation estimator suggested by Barndorff-Nielsen and Shephard (2006) to obtain an estimate of integrated variance that is robust to the presence of jumps. We apply a 'signature plot' similar to that suggested by Andersen, Bollerslev, Diebold and Labys (2001) to identify a sampling frequency at which realized bipower variation appears to be nearly unbiased for most of the stocks in our sample. We then examine the statistical properties of returns standardized by the square root of realized bipower variation.

Accounting for jumps in estimation of the daily integrated variance of returns yields returns that are much closer to a standard normal distribution relative to returns standardized by realized variance. Indeed, we are unable to reject the null hypothesis of normality using standard and robust moment-based tests for many of the stocks in our sample. Our results are robust to alternative sampling frequencies so long as these sampling frequencies are not finer than 4 or 5 minutes, at which point microstructure effects begin to severely contaminate our realized bipower variation estimates.

Our results provide useful empirical guidance regarding appropriate continuous time models for asset prices. On the one hand, the results suggest that, at least for the US stocks we consider, jumps are important and should not be ignored in constructing an appropriate model for prices. On the other hand, we find that a Brownian semimartingale model augmented by a finite activity jump component, where the volatility process evolves independently of innovations in prices, does not exhibit obvious signs of misspecification for most of our stocks.

The empirical findings in this paper contrast with those of Thomakos and Wang (2003) who test for normality of standardized returns for futures contracts on the Deutsche Mark, the S&P 500 index, US Bonds and the Eurodollar and generally conclude that the null hypothesis of standard normal returns cannot be rejected in those markets. Our study is also related to recent work by Ait-Sahalia (2002, 2004). Ait-Sahalia (2002) proposes a method for distinguishing between a Markov diffusion processes and a Markov processes with jumps that is based on a fundamental positivity requirement for the transition function of any Markov diffusion process. This approach does not require high-frequency returns. Empirically, Ait-Sahalia (2002) finds that the transition density extracted from implied volatilities on S&P 500 options is incompatible with an underlying diffusion model, i.e., jumps are needed. It is reassuring that the conclusion that jumps should be included in stock price models obtains under both our approach using high-frequency data and that of Ait-Sahalia (2002).[3] Referring to the Ait-Sahalia's research, Granger (2005) remarks "If further evidence for [jump diffusions] is accumulated, the majority of current financial theory will probably have to be rewritten with "jump diffusion" replacing "diffusion" and with some consequent changes in theorems and results." (p. 18) This paper finds precisely this, in the sense that jump diffusions must be entertained.

The paper is organized as follows. In Section 2 we briefly review the relevant theory to obtain testable implications for standardized returns from Brownian semimartingale and Brownian semimartingale models with jumps. Section 3 discusses methods for measuring variation in prices using high-frequency data that permit us to conduct feasible tests. Section 4 describes the data and our approach to constructing high frequency intraday stock returns. Section 5 considers whether daily returns standardized by the square return of realized variance are Gaussian. Section 6 discusses our approach to extracting the jump component from daily return variation and presents empirical results for tests of whether returns, standardized by the square root of bipower variation, are standard normal. Section 7 concludes.

2. Theory

We consider a log stock price process p_t that evolves continuously in time. Let $r_{t,\Delta} = p_t - p_{t-\Delta}$ denote the stock's Δ-period return ending at time t. We will consider intraday

[3]Ait-Sahalia (2004) shows that it is possible to distinguish between jumps and diffusion models even when jumps are small and occur infinitely often in finite time interval.

returns and variance measures, and these are indexed by n for each trading day. In a frictionless market with no arbitrage opportunities p_t must obey a semimartingale process on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, as detailed by Back (1991). Such a process may be decomposed as follows

$$p_t = A_t + M_t$$

where A_t is a (local) finite variation process and M_t is a (local) martingale. The most familiar and most frequently applied process in the class of semimartingales is the Brownian semimartingale

$$p_t = \int_0^t \mu(s)ds + \int_0^t \sigma(s)\, dW(s) \tag{1}$$

where $\int_0^t u(s)ds$ is a local finite variation process, σ is a cadlag volatility process, $dW(s)$ represents increments to a Brownian motion $W(t)$. We note that the process in (1) has continuous sample paths so that no jumps in stock prices are permitted.

Under the additional assumption that the conditional mean and volatility processes $\mu(s)$ and $\sigma(s)$ are independent of the innovation process $W(s)$ over the period $[t-\Delta, t]$, Andersen, Bollerslev, Diebold and Labys (2003) show that

$$r_{t,\Delta}\, |\mathcal{F}_{t-\Delta,t} \sim N\left(\int_{t-\Delta}^t \mu(s)ds, \int_{t-\Delta}^t \sigma^2(s)ds\right) \tag{2}$$

where $\mathcal{F}_{t-\Delta,t}$ is shorthand for the σ-field generated by $\{\mu(s), \sigma(s)\}$ over the interval $[t-\Delta, t]$.[4] If we additionally rule out time variation in the conditional mean of returns, the distributional result in (2) implies that discretely sampled returns are governed by a normal mixture, where the mixture is governed by the daily integrated variance of returns. Conditional mean variation is empirically an order of magnitude smaller than return variation at short horizons, so ignoring potential time variation in the conditional mean should be innocuous in practice. As ABDL (2003) point out, the distributional result in (2) conditions on *ex post* sample path realizations of $\mu(s)$ and $\sigma(s)$.

If the price process follows (1) with $\sigma(s)$ and $\mu(s)$ independent of $W(s)$ and assuming it were possible to observe the intraday integrated variance, then intraday returns standardized by the square root of intradaily integrated variance should follow a standard normal distribution. We might conduct a statistical test to determine whether this implication

[4]Henceforth the authors Andersen, Bollserslev, Diebold and Labys will be referenced as "ABDL," while Andersen, Bollerslev, Diebold and Ebens will be similarly referenced as "ABDE."

appears to be empirically valid. Rejection of the null hypothesis that standardized returns are Gaussian is thus implicitly a rejection of the model (1) with the additional independence restrictions.

The assumptions underlying (2) rule out features of stock prices that may be important empirically. It is possible that stock prices exhibit jumps resulting from the sudden arrival of firm-specific or macroeconomic news. Further, the assumption that both the mean and volatility process are independent of the increment to returns rules out the so-called *leverage effect*, whereby innovations to returns and volatility are contemporaneously correlated. As an alternative to (1), we also consider a continuous-time semimartingale jump diffusion process:

$$p_t = \int_0^t \mu(s)ds + \int_0^t \sigma(s)\, dW(s) + \sum_{j=1}^{N(t)} \kappa_j \tag{3}$$

where N is a counting process, κ_j are the associated nonzero jump increments and all other components are as in (1). The process in (3) may be characterized as a Brownian semimartingale with finite jump activity. It is worth emphasizing that the process (3) is quite general. The drift component $\mu(s)$ is simply assumed to be a locally finite variation process, thus a time-varying drift is acceptable so long as the finite variation condition is satisfied. All that is required of the volatility process $\sigma(s)$ is that it be cadlag. Again, much flexibility is permitted, such as multi-factor volatility processes as studied by Chernov, Gallant, Ghysels and Tauchen (2003). Finally, little is specified regarding the jump component save that an infinite number of jumps in a 'small' period is forbidden. In particular, the jump intensity and jump size may evolve over time.

Suppose that both the integrated variance process and jump process $\sum_{j=1}^{N(t)} \kappa_j$ were observable. We could then construct the following adjusted price process and associated intradaily return series:

$$p_t^* = p_t - \sum_{j=1}^{N(t)} \kappa_j \tag{4}$$

$$r_n^* \equiv p_{n^C}^* - p_{n^O}^* = r_n - \sum_{j=N(n^O)+1}^{N(n^C)} \kappa_j. \tag{5}$$

where n^O and n^C denote the times corresponding to the market open and close on trading day n. The series in (5) is simply the daily returns less the return component arising from jumps. The process p_t^* is a Brownian semimartingale with continuous sample paths. If we maintain the assumption that the conditional mean and volatility processes $\mu(s)$ and $\sigma(s)$ are independent of the innovation process and we again ignore time variation in the conditional mean, then daily returns r_n^*, standardized by the square root of daily integrated variance should follow a standard normal distribution. We could again test whether this implication holds empirically. Given the generality of the process (3), the most empirically relevant restriction on the price process is the assumption of independence between $\{\sigma(s), \mu(s)\}$ and $W(s)$, which is indeed restrictive. This assumption rules out the leverage effect and more generally rules out any type of dependence between evolution in the mean and variance process and evolution of the Brownian motion governing return increments. This includes, for example, volatility "feedback" effects in which future increments to returns depend on the past evolution of volatility.

Of course, since neither the integrated variance series nor the jumps series in the case of the process (3) are observed, the tests described above are infeasible. Making use of recent econometric results that rely on high-frequency intraday price data, however, we will be able to construct feasible versions of these tests. The following subsections briefly summarize recent results that will permit us to decompose the total variation in prices into integrated variance and a variation attributable to jumps in asset prices.

3. Constructing Feasible Tests

We begin with the fundamental notion of the *quadratic variation* (QV) of the price process p_t, defined as

$$[p]_t \equiv p \lim_{m\to\infty} \sum_{i=1}^{t_i \leq t} (p_{t_i} - p_{t_{i-1}})^2 \tag{6}$$

for any (deterministic) sequence of partitions $0 = t_0 < t_1 < ... < t_m = t$ with $\sup_i\{t_{i+1} - t_i\} \to 0$ for $m \to \infty$. Under the assumption that p_t is a semimartingale the quadratic variance $[p]_t$ is assured to exist. For the process in (3), the quadratic variation of p_t may be decomposed as

$$[p]_t = \int_0^t \sigma^2(s)ds + \sum_{j=1}^{N(t)} \kappa_j^2. \tag{7}$$

where $\int_0^t \sigma^2(s)ds$ is the integrated variance of the price process and $\sum_{j=1}^{N(t)} \kappa_j^2$ is simply the sum of squared jumps. The increment in the quadratic variation process during trading day

n is given by

$$[p]_{n^C} - [p]_{n^O} = \int_{n^O}^{n^C} \sigma^2(s)ds + \sum_{j=N(n^O)+1}^{N(n^C)} \kappa_j^2. \tag{8}$$

Intuitively, the day n increment in the quadratic variation of the price process captures the total intraday variation in the price process, which may be decomposed into the sum of the integrated variance for the n^{th} trading day plus the sum of squared jumps occurring on the n^{th} day.

3.1 Realized Variance

Econometric measurement of the increments in quadratic variation is typically based on the *realized variance*. The standard definition of the realized variance on day n is obtained by dividing the trading day into m subperiods of length $\Delta = 1/m$ and then summing the squared returns for each subperiod,

$$\mathrm{RV}_n^m = \sum_{i=0}^{m-1} r^2_{n^C - i\Delta, \Delta}. \tag{9}$$

ABDL (2001) and Barndorff-Nielsen and Shephard (2001, 2002) show that under weak regularity conditions, the theory of quadratic variation implies that $\mathrm{RV}_n^m - \mathrm{IV}_n \to 0$ almost surely as $\Delta \to 0$. In the absence of jumps, the decomposition in (7) implies that the quadratic variation is simply equal to the integrated variance. Thus, by using high-frequency intradaily returns, we can potentially construct nonparametric estimates of the integrated variance that are consistent and vastly more efficient than estimates obtained using daily squared returns. Realized variance sampled at sufficiently high frequencies (e.g., every 5 seconds) theoretically yields an effectively error-free measure of quadratic variation. The realized variance estimator, along with the availability of high-frequency intraday returns, thus permits a feasible test of the normal mixture hypothesis for standardized daily returns.

3.2 Bipower Variation

Related to the quadratic variation process is the *bipower variation process*, defined as[5]

$$\{p\}_t = \text{p-}\lim_{\delta \to 0} \sum_{i=2}^{\lfloor t/\delta \rfloor} |p_{i-1}| \ |p_i|. \tag{10}$$

[5] Bipower variation may be defined more generally (see, e.g., Barndorff-Nielsen and Shephard (2005)). However, we will not need this additional generality in this paper.

Unlike the quadratic variation process, which is assured to exist for semimartingales, the bipower variation process is not guaranteed to exist. The limit in (10) does exist, however, when log prices follow the process (3). Barndorff-Nielsen and Shephard (2006, hereafter "BNS") show that, if log prices follow (3) with mild additional restrictions and under the additional assumption that σ is independent of $W(s)$ (this rules out the leverage effect), then

$$\{p\}_t \to \frac{2}{\pi} \int_0^t \sigma^2(s) ds.$$

The crucial implication of this result is that $\frac{\pi}{2}\{p\}_t$ and quadratic variance are identical when there are no jumps (i.e., under process (1)), but differ precisely when jumps occur. Again dividing the trading day into m subperiods of length $1/\Delta$, BNS show that a consistent estimator for the daily increment in the bipower variation process is

$$\sum_{i=2}^{m} \left| r_{n^C - i\Delta, \Delta} \right| \left| r_{n^C - (i+1)\Delta, \Delta} \right|.$$

Multiplying the above estimator by $\pi/2$, we obtain a consistent estimator for quadratic variance in the presence of jumps. Following the literature, we will define *realized bipower variation* (BV) as

$$\text{BV}_n^m = \frac{\pi}{2} \frac{m}{m-1} \sum_{i=2}^{m} \left| r_{n^C - i\Delta, \Delta} \right| \left| r_{n^C - (i+1)\Delta, \Delta} \right| \tag{11}$$

where the $\frac{m}{m-1}$ term is a finite sample correction that delivers unbiased estimates in the Brownian motion case (see BNS).

With sufficiently finely sampled data, BV provides a highly accurate estimate of the daily integrated variance in returns in the presence of jumps. By using the BV series to standardize returns, we make progress toward a feasible test of the normal mixture hypothesis when returns exhibit jumps. The remaining obstacle is the need to construct a feasible estimator for the jump component in returns, $\sum_{j=N(n^O)+1}^{N(n^C)} \kappa_j$. Extracting the jump component from returns is a difficult endeavor. To appreciate this, note that while a comparison of the difference $RV_n^m - BV_n^m$ yields a consistent estimate of the daily contribution of jumps to total return variation, $\sum_{j=N(n^O)+1}^{N(n^C)} \kappa_j^2$, such a comparison does not directly provide information regarding the number, sign, or magnitude of jumps in asset prices. We defer further consideration of this issue until Section 6 of the paper where we discuss the specifics of our approach to testing whether standardized returns 'adjusted' for jumps are Gaussian.

3.3 Microstructure effects and realized variance

In practice, realized variances constructed according to equation (9) are often biased because the true price process is unobservable and the variances must be estimated from observed prices. For returns constructed from transaction prices, serial correlation is induced by bid-ask bounce, price discreteness, and price reporting errors. For returns constructed from bid-ask midpoints, serial correlation is induced by the second of these effects, as well as the economics driving dealer revisions of their bid-ask quotes. Using either approach, the serial correlation tends to be more pronounced with more frequent sampling (i.e., increasing m), which suggests that the bias can potentially offset the efficiency gains associated with more frequent sampling.[6]

Given the trade-off between bias and efficiency gains, it has become standard practice to construct realized variance using returns sampled at a five-minute frequency (e.g., see Andersen and Bollerslev (1997)). More formally, Bandi and Russell (2004) derive the optimal sampling frequency under a mean squared error criterion. Other researchers use techniques to reduce the effects of the measurement error.[7] ABDE (2001), for example, fit an MA(1) model to observed returns and then use the residuals to construct realized variance. ABDL (2001) use linear interpolation to estimate the price at each time increment and then use the returns based on these "smoothed" prices to construct realized variance. However, using either of these techniques, returns still typically exhibit serial correlation, and thus we must still limit the sampling frequency in order to prevent bias.

We adopt another approach which directly accounts for the serial correlation in returns when constructing our estimator of the variance. Specifically, we use a Newey-West (1987) estimator of realized variance,

$$\mathrm{RV(NW)}_n^m = \sum_{i=0}^{m-1} r^2_{n^C - i\Delta,\Delta} + 2\sum_{j=1}^{q_m}\left(1 - \frac{j}{q_m+1}\right)\sum_{i=0}^{m-j-1} r_{n^C - i\Delta,\Delta}\, r_{n^C-(i+j)\Delta,\Delta}, \qquad (12)$$

where q_m denotes the lag length captured by the covariance terms. The advantage of this approach is that the estimator removes the effects of serial correlation on the variance, potentially allowing us to sample returns at a higher frequency and incorporating relevant information that might otherwise be lost. We evaluate different choices of the sampling frequency (m) and the lag length (q_m) in Section 5.

The estimator in (12) is motivated by bias correction results obtained by Hansen and

[6]See, e.g., Andreou and Ghysels (2002) for some numerical estimates of the bias.

[7]A number of recent papers examine microstructure noise and realized variance estimation, including Barndorff-Nielsen, Hansen, Lunde and Shephard (2004), Oomen (2002, 2004), Hansen and Lunde (2006) and Zhang, Mykland and Ait-Sahalia (2004).

Lunde (2006) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2004). Hansen and Lunde (2006) propose a different weighting scheme in equation (12) which delivers conditionally unbiased estimates. However, their estimator can generate negative variance estimates. In our sample, the number of negative estimates per stock ranges from 42 with a 30-second sampling frequency, to 25 with a 5-minute sampling frequency, to 135 with a 60-minute sampling frequency. The estimator in (12) is guaranteed to be positive. This comes at a cost, however, since this estimator is not, strictly speaking, conditionally unbiased. In practice, however, the estimator affords the potential of a nearly unbiased realized variance estimator that is guaranteed to be positive. We confirm the nearly unbiased behavior of the NW estimator empirically in Section 5. The NW estimator thus provides an alternative to the 'raw' realized variance estimator for conducting feasible tests for the normality of standardized returns.

In Section 5 of the paper we explore the extent to which microstructure effects bias standard realized variance estimates and the ability of the Newey-West estimator of realized variance to correct the bias. Before moving on to describe the data and our construction of high-frequency returns, we note that microstructure effects might also affect the BV estimator. This issue has received fairly limited attention in the literature. Huang and Tauchen (2005) and BNS examine bias in realized bipower variation under different models for microstructure noise. Both studies recommend using a "skipped" BV estimator to mitigate potential bias from microstructure noise. Such a skipped estimator takes the form

$$\mathrm{BV}_n^{m,j} = \frac{\pi}{2}\frac{m}{m-j}\sum_{i=j+1}^{m} \left|r_{n^C-i\Delta,\Delta}\right| \left|r_{n^C-(i+j)\Delta,\Delta}\right| \tag{13}$$

where j indicates the degree of skip. BNS and Huang and Tauchen (2005) suggest that a single skip may work well in practice. BNS also point out that the BV estimate appears to fall as the sampling frequency becomes less fine. Fleming and Paye (2005) empirically examine the sensitivity between sampling frequency and skipped and un-skipped versions of BV. They conclude that BV estimates, like RV estimates, are subject to bias at high sampling frequencies and suggest using an analog of the signature plot of ABDL (2000) to determine an appropriate sampling frequency for unbiased estimation of BV. We return to this issue in Section 6 of the paper where we discuss the empirical details of our approach to detecting and extracting jumps in stock prices.

4. Data

The data consist of intradaily prices for the 20 stocks in the Major Market Index (MMI).[8] We obtain these data from the Trade and Quote (TAQ) database distributed by the New York Stock Exchange (NYSE). The sample period is January 4, 1993 to December 31, 2003, a period of 2,771 trading days.[9]

We construct a database of intradaily transaction prices from the TAQ consolidated trade (CT) database. We include records for trades executed on any exchange, but we exclude records that have an out-of-sequence time, a price of zero, a TAQ correction code greater than two (i.e., errors and corrections), or a TAQ condition code (i.e., nonstandard settlement). In addition, we apply two screens to eliminate obvious price reporting errors. First, we eliminate all records with a reported price more than 20 percent greater or less than the previous transaction price. Second, we identify all records that imply a price change greater than two percent in magnitude, which are followed by a price reversal greater than two percent in magnitude. We eliminate the record if the price change is more than two times greater than the next largest price change on that day, or if the "suspicious" price is outside the day's high-low range by more than the next largest price change on that day.

4.1 Construction of the returns

On most days, the trading day begins at 9:30am EST and ends at 4:00pm EST, a period of 390 minutes. We divide this period into m intervals to construct returns with a sampling frequency of $390/m$ minutes. We consider a range of values for m, from $m = 780$ (i.e., 30-second intervals) to $m = 6.5$ (i.e., 60-minute intervals). If m is an integer (e.g., $m = 13$), then the trading day divides evenly into m intervals, each with a length of $390/m$ minutes. If m is not an integer (e.g., $m = 6.5$), then we construct $\text{ceil}(m)$ intervals, with the first one only $\text{mod}(390, \text{int}(m))$ minutes in length, and the others $390/m$ minutes in length.[10]

We identify the prices at the endpoints of each of these intervals using the price database from the previous section. We start with the first price in the database for the particular day. We treat this as the starting price for the interval in which the observation occurred. We

[8] The firms in the MMI are American Express (AXP), AT&T (T), ChevronTexaco (CVX), Coca-Cola (KO), Disney (DIS), Dow Chemical (DOW), DuPont (DD), Eastman Kodak (EK), Exxon-Mobil (XOM), General Electric (GE), General Motors (GM), International Business Machines (IBM), International Paper (IP), Johnson & Johnson (JNJ), McDonald's (MCD), Merck (MRK), 3M (MMM), Philip Morris (MO), Procter and Gamble (PG), and Sears (S).

[9] Philip Morris has one less trading day than the other firms because its stock did not open on May 25, 1994 in advance of a board meeting regarding a proposal to split the firm's food and tobacco businesses.

[10] $\text{ceil}(m)$ denotes the smallest integer greater than or equal to m, $\text{int}(m)$ denotes the smallest integer less than or equal to m, and $\text{mod}(a, b)$ denotes the remainder of a divided by b.

then estimate the price at the end of this interval using linear interpolation (Andersen and Bollerslev (1997)).[11] We identify the first price before (p_{t_i}) and the first price after ($p_{t_{i+1}}$) the end of the interval (at time τ), and we compute $p_\tau = p_{t_i} + (p_{t_{i+1}} - p_{t_i})(\tau - t_i)/(t_{i+1} - t_i)$. If one or more observations occur exactly at the end of the interval, then we average the prices for all of these observations rather than interpolating. We use this same procedure to estimate the price at the end of each of the remaining intervals throughout the day, except for the last interval, in which we simply use the day's last price. We then take the first differences of the logs of these prices to estimate the $390/m$-minute intraday returns.

4.2 Characteristics of Returns

Table 1 displays basic summary statistics for each of the stocks in the MMI, including the average number of trades, the average intraday squared return and the annualized volatility of intraday returns. On average, there are over 2,200 transactions per day per stock during our sample period. GE has the greatest number of transactions, 5,533 per day on average, and DOW has the least, 965 per day on average. Sears has the highest volatility, with an average daily squared open-to-close return of 4.36 percent (or 33.2 percent annualized volatility), and XOM has the lowest volatility, with an average daily squared open-to-close return of 1.80 percent (or 21.3 percent annualized volatility). Table 1 also reports the average time in seconds to the 5-minute grid, defined as the average of the time distance of the nearest trade across all gridpoints based on our partition of all trading days into 5-minute subintervals. Lower average times to gridpoint are indicative of more active markets. When the average time to gridpoint is high relative to the sampling frequency (e.g., relative to 300 seconds for the results in Table 1) then microstructure effects are likely to be substantial.

Table 2 reports the serial correlation of the intraday returns by sampling frequency, averaged across the 20 stocks. As expected, the first-order serial correlation is typically negative (-0.07 for five-minute returns), and the magnitude increases with the sampling frequency (-0.15 for 30-second returns). As noted earlier, these effects can arise from bid-ask bounce, price discreteness, and/or reporting errors, and for each of these causes the effects should be more pronounced at finer sampling frequencies.

[11] We find that using linear interpolation rather than the last price in each interval substantially reduces the serial correlation in returns at high sampling frequencies and leads to less biased realized variance estimates. This is true even if we use an MA(1) model to filter the returns (e.g., ABDE (2001)). Moreover, although Hansen and Lunde (2006) show that $\mathrm{RV}_n^m \xrightarrow{p} 0$ as $m \to \infty$ when using the interpolation method, we find no evidence of this effect in our sample using sampling frequencies as fine as 30 seconds. As a robustness check we also conducted the empirical analysis using prices constructed using the last-tick method as opposed to interpolation. The main qualitiative results are very similar and are available from the authors upon request.

5. Are Returns Standardized by Realized Volatility Gaussian?

While there is a great deal of emerging literature that examines microstructure bias and realized variance estimation, a standard applied practice entails sampling returns at a 5-minute frequency. As an exploratory analysis, we construct realized variance estimates at the 5-minute frequency and examine the properties of the realized variances and daily returns standardized by the realized volatility. Table 3 reports summary statistics for the realized volatility and log realized volatility for each of the MMI firms. Realized volatility exhibits positive skewness and heavy excess kurtosis. Log realized volatility, on the other hand, does not display marked skewness or leptokurtotic behavior.

Table 4 presents results for a battery of statistical tests for the normal distribution based on the standardized returns. In this paper we focus on moment-based tests for normality. We report p-values for the Jarque-Bera test as well as for separate tests based on the estimated skewness and kurtosis. Tests based on skewness and kurtosis statistics may be quite sensitive to outliers. For this reason, we also report robustified versions of these tests due to Brys, Hubert and Struyf (2004b). The appendix to the paper discusses the construction of these robust tests, which sacrifice some power relative to the standard tests but are less sensitive to outliers. Table 4 presents the summary statistics and results for each of the tests for all of our MMI firms. The appendix contains additional details regarding the tests for normality reported in the table.

While there is some variation both across stocks and across tests, the results in Table 4 appear to offer a reasonable amount of support for the simple normal mixture hypothesis. The Jarque-Bera tests fail to reject at the 5% level for 7 out of 20 stocks and fails to reject at the 1% level for 17 out of 20 stocks. The robust test statistics fail to reject for most of the stocks. Our results in Tables 3 and 4 broadly replicate findings reported in ABDE(2001), who find that standardized daily stock returns are 'approximately Gaussian.' ABDL(2000) and ABDL(2003) report similar findings for exchange rate returns.

It is tempting to simply conclude that the normal mixture hypothesis is empirically supported based on these test results. On the other hand, the tests were based on a 5-minute sampling frequency which was somewhat arbitrarily selected to accord with common practice. It is possible that microstructure effects exist at this sampling frequency, which may impact statistical inference. Consequently, we next explore the robustness of our test results to the choice of sampling frequency.

5.1 Microstructure bias and the kurtosis of standardized returns

We construct the trading-day realized variances given by equations (9) and (12) for each stock, for each day in the sample, and we evaluate a range of sampling frequencies, from $m = 780$ to $m = 6.5$. We refer to estimates based on (9) and (12) as the RAW and NW estimates, respectively. For the NW estimates, we also evaluate a range of window lengths (w) for the autocovariance terms: 15, 30, and 60 minutes. To implement a given window length, we set $q_m = \text{ceil}(\frac{w}{390/m})$.[12] We refer to these as the NW15, NW30, and NW60 estimates. Figure 1 shows the average realized variances for each sampling frequency, averaged across all of our MMI firms. The figure is similar to the volatility signature plot proposed by ABDL (2000) to identify the 'optimal' sampling frequency.[13] The idea is to detect the frequency at which microstructure effects begin to bias the estimates. ABDL (2000) recommend selecting the finest sampling frequency for which microstructure effects do not appear to bias estimates. The tick mark labeled DSR on the y-axis shows the average daily squared open-to-close return.

The highest sampling frequency at which realized variance estimates appear to be unbiased is substantially greater than 5 minutes for the RAW estimator. Sampling more frequently, the realized variance estimates are upward biased due to the negative serial correlation in returns, and sampling less frequently (although not biased) entails greater sampling error. At a five-minute frequency, the estimates are on average more than 13 percent greater than the average daily squared return.[14] Also note that the bias increases sharply at the highest sampling frequencies.

The NW estimates, on the other hand, are not very sensitive to the sampling frequency, and they show little evidence of bias. The NW30 and NW60 estimates are especially stable, and for all of the sampling frequencies, they are within 2 percent of the average daily squared return. There is some evidence that increasing the window width (w) beyond 15 minutes reduces bias. The NW30 and NW60 estimates are on average smaller than the NW15 estimates and closer to the DSR, however, there is little noticeable difference in increasing w from 30 to 60 minutes.

[12] As in Hansen and Lunde (2006), we increase q_m when increasing the sampling frequency, m. The intuition is that measurement errors take a certain period of time to die out (e.g., 15 minutes), and q_m should cover this period, independent of the sampling frequency.

[13] To conserve space, we do not present separate figures for each stock; however, such figures are available upon request and illustrate that the individual stock plots share the features of the plot for the average across stocks.

[14] We find that using MA(1)-filtered returns to construct realized variance (ABDL (2001)) does not remove much of the bias. At the 5-minute frequency, the average MA(1) realized variance is less than one percent lower than the average RAW estimate and, at the 30-second frequency, the difference is less than five percent.

Figure 1 also presents the standard deviations of the RAW and NW estimates, averaged across all MMI firms. For all choices of w, the standard deviations decrease with the sampling frequency. This suggests, to the extent that the estimates remain conditionally unbiased, that the efficiency gains from more frequent sampling lead to more optimal estimates. Also note that the standard deviations are increasing in w. This is not surprising because including additional (unnecessary) covariance terms reduces efficiency. Increasing w from 15 to 30 minutes may be optimal to achieve a reduction in bias, but increasing to 60 minutes achieves little bias reduction and entails a noticeable loss in efficiency.[15]

We also use signature plots to examine the sensitivity of estimates of the skewness and kurtosis of standardized daily returns to the sampling frequency. Figure 2 (Panel A) plots the average skewness across the MMI firms for standardized returns, $r_n/\sqrt{RV_n}$, as a function of the sampling frequency for the the RAW realized variance estimator and the NW variants. The skewness estimates in Figure 2 are reasonably flat as a function of the sampling frequency for all realized variance estimators. To the extent that microstructure noise biases the RAW realized variance estimates, there does not appear to be a corresponding bias induced in the skewness of standardized daily returns.

The situation is much different for a similar plot for kurtosis estimates, presented in Panel B of Figure 2. It is clear that the upward bias in the RAW realized variance estimator induces a corresponding upward bias in the kurtosis estimates. The signature plot of kurtosis under the RAW estimator for most stocks may be described as a downward sloping curve starting at roughly 3.5 at the highest sampling frequency and ultimately flattening out at a kurtosis in the range 2.3 - 2.5. The NW estimators, on the other hand, exhibit flat kurtosis signature plots, mostly in the 2.3 - 2.7 range. It is notable that at the 5-minute sampling frequency the kurtosis under the RAW estimator is close to 3 in many cases. The signature plots reveal that not all microstructure bias has been purged at this sampling frequency and these estimates of the kurtosis of standardized returns remain upward biased. Table 5 presents tests for normality for returns standardized by the square root of the NW-30 estimate of realized variance at the 30-second sampling frequency. The results differ markedly from those displayed in Table 4. The Jarque-Bera test rejects for all stocks at the 99% level. The separate moment tests reveal that the rejection is largely driven by the platykurtotic behavior of these standardized returns. We continue to reject the null in most cases under the two robust tests that consider tail-weight (RJB1 and RJB2). In summary, once an unbiased realized variance estimator is used to standardize returns, it is

[15] We also investigating using bid-ask quotes to construct realized variance. However, we find that the variance of the estimates is approximately the same as that using transaction prices.

clear that, from a classical hypothesis testing viewpoint, the null of normality is not difficult to reject.

5.2 What drives the bias in kurtosis?

The signature plots in Figures 1 and 2 illustrate that the upward bias in realized volatility at high sampling frequencies yields a corresponding upward bias in the kurtosis of standardized returns. Consequently, reliance on RAW realized volatility estimates at the standard 5-minute sampling frequency leads to a spurious finding that standardized returns are normally distributed for the stocks we examine. We now explore the mechanism by which the upward bias in realized volatility translates into an upward bias in the kurtosis of standardized returns.

There is reason to suspect that the bias in RAW realized variance estimates varies over time as market conditions fluctuate. Fundamentally, this bias is induced by negative autocovariances in high-frequency returns related to microstructure effects such as bid-ask bounce. The nature of these microstructure effects presumably varies over different trading days according to market conditions. For example, the extent to which microstructure effects contaminate high-frequency returns may depend on whether an unanticipated firm-specific or macroeconomic announcement occurs on the trading day in question. Suppose that microstructure effects are weaker on days characterized by large absolute returns. Then the RAW realized variance will be less upward biased on large (absolute) return days relative to other trading days. Consequently, when returns are standardized by RAW realized volatility, the relatively smaller upward bias in realized volatility on large return days will artificially inflate the tails of the unconditional distribution of standardized returns, leading to upward bias in kurtosis estimates for the standardized returns. We now present empirical evidence consistent with this hypothesized mechanism for the stocks we examine.

We designated, on a stock-specific basis, trading days when the absolute daily return was in excess of 2 percent as large return days. The number of large return days varies from 763 for Sears, to 316 for ExxonMobil. Table 6 presents summary statistics for daily returns on large return days. Comparing the results in Table 6 to those in Table 1, it is clear that large return days are accompanied on average by higher trading volume and higher volatility. The average time to the gridpoint is substantially lower for many of the stocks on days with large returns. Table 7 presents the average of autocorrelation coefficients across the MMI stocks at various sampling frequencies for large return days. When compared to the analogous results in Table 2 for all daily returns, it is apparent that returns on large return days exhibit much less negative serial correlation at high sampling frequencies, consistent with the notion that microstructure effects are weaker on these trading days.[16]

Figure 3 displays the average realized variances across the 20 MMI stocks as a function of the sampling frequency on large return days. Clearly, the volatility signature plot conditional on a large absolute return exhibits much less evidence of upward bias due to microstructure effects relative to the corresponding signature plots for all daily returns. These results are consistent with the notion that microstructure effects are weaker on large return days and that this phenomenon drives the transmission of upward bias in realized variance estimates to upward bias in the kurtosis of standardized returns.

6. Do jumps account for the non-normality of standardized returns?

The results in the preceding section strongly indicate that daily returns standardized by realized volatility are not Gaussian, and in particular they are platykurtotic. An obvious explanation for non-normality of the standardized returns is the presence of jumps. In this section we describe our approach to determining whether standardized returns are normal once we account for jumps We then discuss the empirical results and comment on alternative approaches and the robustness of our results.

6.1 Methodology

We seek to construct estimates of the series $\{r_n^*/IV_n\}$ where IV_n denotes the integrated volatility on day n, and test whether the resulting series is standard normal as implied by the theory discussed in Section 2. As discussed in Section 3 of the paper, the realized bipower variation provides a consistent estimate of the integrated variance of returns in the presence of jumps. Thus, rather than standardize returns by the realized volatility, we standardize returns using in the square root of the realized bipower variation BV_n^m defined in (11). There remains the choice of sampling frequency, m, on which to base BV_n^m While the theory of bipower variation suggests that we sample as frequently as possible in order to obtain as accurate an approximation to the IV_n as possible, this ignores the possibility that microstructure effects may induce bias in our estimates.

We elect to use bipower variation 'signature plots' similar to those proposed by ABDL (2001) as a practical tool for selecting an appropriate sampling frequency for computing bipower variation estimates. As in the case of realized variance, the 'optimal' sampling

[16]The autocorrelations on these trading days are positive and substantial at sampling frequencies greater than 30 minutes. This is an artifact of our conditioning on large absolute return days. To see this, consider half-day returns. Conditional on a large return day, returns over the first and second half of the day will tend to be positively autocorrelated as, e.g., a positive return in the morning is more likely to be followed by a subsequent positive return given that at close there is greater than a 2% daily return in absolute value.

frequency is the highest frequency at which the BV_n^m estimates appear to be unbiased. In the case of realized variance, the average daily squared return is typically taken as a yardstick against which to measure bias. In the case of realized power variation, BV_n^2 serves as a similar yardstick. This is simply the realized bipower variation computed using half-day returns. At this frequency, microstructure effects are likely to be minimal, particularly for the heavily-traded stocks considered in this paper, and so this estimator provides a benchmark against which to compare estimators at higher frequencies.

Figure 4 presents a signature plot for the average realized bipower variation averaged across the MMI stocks. The figure plots average realized bipower variation for BV_n^m (denoted BV) as well as for 'skipped' versions computed according to (13) for $j \in \{2, 3\}$.The figure illustrates that BV_n^m is substantially upward biased for sampling frequencies higher than 4-5 minutes and that a sampling frequency somewhere between 8-15 minutes is need to achieve approximately unbiased estimates on average across the stocks. Based on the signature plot in Figure 4, we elect to base realized bipower variation on BV_n^{10}, i.e., on a sampling frequency of 10 minutes. At this frequency there appears to be minimal bias for many of the stocks we consider.[17]

The final obstacle to a feasible test is construction of r_n^*, the daily return adjusted to remove the effect of jumps. As noted in Section 3, this is a nontrivial problem. Detection of jumps may be based on a comparison of RV_n^m and BV_n^m and statistical tests for the presence of jumps have been proposed by BNS (2006) and Huang and Tauchen (2005). However, even though it is possible to detect the occurrence of jumps on a given trading day, such tests provide little direct information regarding the number of jumps or the direction and magnitude of jumps. Our approach is guided by the desire to obtain an estimate of r_n^* that is unbiased irrespective of whether or not jumps occur on a given trading day. If no jumps occur, then r_n is equivalent to r_n^* Suppose that one or more jumps occur. If the daily return provides minimal information regarding the sign and magnitude of the jumps, then r_n should be an approximately conditionally unbiased estimator for r_n^*. Our approach, then, is to simply take $r_n^* = r_n$ to obtain a feasible test. Our new standardized series thus differs from the original standardized series only through the use of a measure of integrated variance that is robust to the presence of jumps.

[17] We also examined firm-by-firm signature plots. There is some variation among the firms regarding the time at which realized bipower variance estimates appear to stabalize. A 10 minute sampling frequency appears to offer reasonably unbiased estimates for most of our firms and for simplicity we apply this sampling frequency in all cases. Robustness of our results to the sampling frequency is discussed in Section 6 of the paper. Additionally, we found that using the skipped versions of the estimator often only partially mitigates the bias at high sampling frequencies and so we elect to use the unskipped version at the 10-minute frequency in this paper.

We must acknowledge that in some cases the daily return may provide valuable information regarding the sign and magnitude of the jump component in the daily return. This may be the case, for example, when a single jump of very large magnitude occurs on a particular trading day. Tauchen and Zhou (2005) attempt to exploit such information and extract the jump component in daily returns. We discuss the robustness of our findings to basing standardized returns on their approach later in the paper.

6.2 Empirical Results

Table 8 presents summary statistics and tests for normality for the series $\{r_n/\sqrt{BV_n^{10}}\}$. The summary statistics illustrate that the kurtosis for many stocks is quite close to the appropriate value of three under the null. For most of the stocks, the variance, skewness and kurtosis are similar to the theoretical equivalents for the standard normal. The returns do appear to slightly positively skewed. Turning to the tests for normality, the Jarque-Bera test fails to reject the null at the 5% level for 11 of 20 stocks and fails to reject at the 1% level for 16 out of 20 stocks. Tests based on the separate moments also frequently fail to reject. The robust tests yield qualitatively similar results, although the decision to accept or reject the null hypothesis at conventional levels sometimes varies between the standard and robust versions of the Jarque-Bera test. Broadly speaking, these results illustrate that while daily returns standardized by realized volatility appear to be platykurtotic, the same returns standardized by a volatility measure that is robust to jumps *do* appear to be approximately Gaussian.

When we examine the separate moment tests, the null if rejected far more frequently for the skewness test than for the kurtosis test. For example, the skewness test rejects for 12 out of 20 stocks at the 5% level while the kurtosis test rejects for only 3 out of 20 stocks. Further, in those instances where we do see emphatic rejections of the null hypothesis under the Jarque-Bera test, for GE, GM, KO and T, we see that in 3 out of 4 of these cases the rejection is driven primarily by the significantly positive skewness of standardized returns.

6.3 Robustness

The results in Table 8 are based on a sampling frequency of 10 minutes, which we selected based on our signature plots for realized bipower variation. Figure 5 explores the sensitivity of our moment estimates to the choice of sampling frequency. In Figure 5 we present the skewness and excess kurtosis estimates for the standardized returns $\{r_n/\sqrt{BV_n^{10}}\}$ as a function of sampling frequency m. For the kurtosis, at very high frequencies the excess kurtosis is substantially greater than zero in most cases; however, as Figure 4 illustrates, realized bipower variation appears to be upward biased in at these sampling frequencies. Just as in the case of bias in realized variance, this bias appears to translate into an upward bias in the kurtosis estimates for standardized returns. For most stocks, the kurtosis estimates stabilize by around 4-5 minutes at a level quite close to 3. The skewness estimates, on the other hand, display little sensitivity to the sampling frequency. Our results, then, depend only on choosing a sampling frequency greater than, say, 4 minutes and are not critically dependent upon our choice of 10 minutes. While we find that a sampling frequency of 8-10 minutes provides less biased realized bipower variation estimates relative to higher sampling frequencies, we have computed summary statistics and results for our tests of normality using a 5-minute sampling frequency, which accords with common practice (see, e.g., Andersen, Bollerslev and Diebold (2004), Huang and Tauchen (2005) and Bollerslev, Kretschmer, Pigorsch and Tauchen (2005)). These results are available from the authors on request.[18]

A potentially important feature of our data is the introduction of decimalization on January 29, 2001. To examine the effect of decimalization, we computed both summary statistics and tests for normality over the 733 observations following decimalization as well as over the 733 last observations for the tick-based period. Results are presented in Table 9. The Jarque-Bera test fails to reject in the post-decimalization period for most stocks (14 out of 20) at the 5% level. Of course, there is a loss of power associated with the much smaller sample size (733 observations relative to 2771) so it is not clear whether the fewer rejections derives primarily from lower power or for a closer approximation to the normal following decimalization. Some insight into this question may be derived from comparing the post-decimalization results to the pre-decimalization results computed over a similar sample size of 733. Over the 733 trading days immediately preceding decimalization, we obtain quite similar results. The Jarque-Bera test fails to reject for 15 out of 20 stocks at the 5% level and the moment estimates for the stocks. Further, the moment estimates are fairly similar across the two subsamples. The most substantial difference between the two sets of results appears to be the kurtosis. The kurtosis in the pre-decimalization period is lower on average than that after decimalization and the test based on this moment rejects more frequently during the tick-based period. Overall, it appears that decimalization did not dramatically alter the properties of standardized daily returns, with the caveat that there is some evidence that the kurtosis is closer to the theoretical value over the post-decimalization period.

[18] As the plots in Figure 4 suggest, we continue to find reasonable support for the standard normal distribution even when we sample every 5-minutes.

As a final robustness check, we consider basing r_n^* on $r_n - J_n$, where is an estimate of the jump component in returns on days where (statistically) significant jumps appear to occur. We base our approach on Tauchen and Zhou (2005), who attempt to extract the 'realized jumps' in returns by making two auxiliary assumptions. First, they assume that at most one jump occurs on any given trading day. Second, they assume that the jump size dominates the return for the day, in the sense that the sign of the daily return is equal to the sign of the jump. Under these assumptions, they obtain the following estimator for the daily realized jumps:

$$J_n = sign(r_n) \times \sqrt{(RV_n - BV_n) \times I_n\{\text{test rejects}\}} \quad (14)$$

where $I_n\{\text{test rejects}\}$ is a one on days when the null of no jumps is rejected at the 1% level and is zero otherwise. We apply a statistical test for jumps based on $(RV_n - BV_n)$ suggested by BNS (2006).

Table 10 presents results for standardized returns based $r_n - J_n$, where we continue to use $\sqrt{BV_n^{10}}$ as the standardization factor. These standardized returns are platykurtotic, although they are less platykurtotic relative to the raw returns standardized by the NW realized variance (compare with Table 5). It is notable that the standard deviation for these returns is substantially less than one. This is explained by spurious rejection of the null in the test for jumps. Suppose that no jump occurs but a spurious jump is detected by the test. The return is adjusted by subtracting $sign(r_n) \times \sqrt{RV_n - BV_n}$ which always has the effect of reducing the absolute value of the daily return. Spurious jumps thus bias the standard deviation downward. Note that even for a perfectly sized test, spurious jumps will be detected for 1% of the number of days on which no jumps truly occur on average. If jumps occur on 5% of trading days, so that no jump occurs on the remaining 95% of trading days, then for our sample we would expect roughly 26 spurious jumps to be detected even for a perfectly sized test. Although we reject the null that standardized returns are standard normal under this approach, based on the discussion above it is clear that approaches based on statistical tests for jumps lead to bias in the direction of rejecting a standard normal distribution.[19]

[19] We emphasize that the objectives of this paper differ radically from those in Huang and Tauchen (2005b). Indeed, our approach of setting $r_n^* = r_n$ would be completely useless in their study, since their objective is to extract realized jumps in returns and characterize the jumps.

7. Conclusion

Recent empirical research suggests that jumps occur in asset prices. At the same time, daily returns standardized by 'realized volatility' have been found to be approximately standard normal, a finding consistent with a specification that rules out jumps. This paper resolves these apparently contradictory findings using high-frequency returns data for 20 heavily-traded US stocks. We show that kurtosis estimates based on standardized returns computed using the typical realized variance estimator can be upward biased due to microstructure noise when data are sampled at very high frequencies. Using a realized variance measure that corrects for microstructure bias, we find that standardized returns are in fact platykurtotic and the normal distribution can be statistically rejected.

If jumps are the source of non-normality for standardized returns, then standardized daily returns should follow the standard normal distribution when the standardization is based on realized bipower variation, an estimator for integrated volatility robust to the presence of jumps. This is precisely what we find for many of our stocks. In fact, for roughly half of the stocks we consider our 'adjusted' standardized returns are statistically indistinguishable from the standard normal distribution when confronted with a set of hypothesis tests. Even in instances when we continue to reject the null of normality, the returns standardized by realized bipower variation provide a closer approximation to the normal distribution relative to returns standardized by an unbiased measure of realized variance. This finding is robust to the specific choice of sampling frequency so long as this frequency is greater than 3 or 4 minutes.

The mixture of normals hypothesis places important restrictions on the stock price process. Stock price paths must be continuous and the volatility process must evolve independently of the Brownian motion governing returns. Our results illustrate that the first restriction is violated in the data. Our results also suggest that, at least for a number of the stocks we consider, that the second restriction may be reasonably innocuous. Among other implications, this suggests that the leverage effect is relatively unimportant for our individual stocks.

The analysis in this paper is based on a set of 20 heavily-traded US stocks. It would be interesting to explore whether similar results obtain for other asset classes, including foreign exchange, Treasury instruments, commodities and aggregate indices such as the S&P 500. Some empirical evidence suggests that the leverage effect is stronger for aggregate indices such as the S&P 500 than it is for individual stocks. Ait-Sahalia (2002) rejects a Markov diffusion process for the S&P 500 in favor a process with jumps. What remains to be determined, however, is whether, as in our case, returns standardized by realized bipower variation are standard normal for this index. Another question of interest is whether daily integrated variance appears to follow a normal distribution. ABDL (2000, 2003) provide evidence indicating that realized variance is approximately unconditionally Gaussian. Realized variance, however, may be contaminated by price variation attributable to jumps. It would be interesting to explore whether the log realized bipower variation appears to provide a better approximation to the normal distribution and whether the log realized bipower variation is statistically consistent with the normal distribution. As discussed by Thomakos and Wang (2003) and Bontemps and Meddahi (2003), testing for normality in this setting is more complex due to the presence of serial correlation and perhaps long run dependence in (log) realized bipower variation.

A Appendix

1. Statistical Tests for Normality

This appendix briefly describes the various statistical tests for a standard normal distribution applied in the paper. Several of the tests we apply are standard. The robustified versions of tests for normality, due to Brys, Hubert and Struyf (2004), are less frequently applied in econometrics and for this reason we devote relatively more space to a discussion of these tests. Let $\{X_t\}_1^T$ denote a time series of *iid* observations that, under the null hypothesis, represent a random sample from the standard normal distribution. The sample average and sample standard deviation are $\bar{X} = T^{-1} \sum_{t=1}^{T} X_t$ and $\hat{\sigma} = \{T^{-1} \sum_{t=1}^{T} \left(X_t - \bar{X}\right)^2\}^{1/2}$.

1.1 Moment-based tests for normality

The sample skewness and sample kurtosis may be expressed as functions of the sample mean and sample standard deviation. These statistics are

$$\begin{aligned} \widehat{\mathcal{S}}_T &= \frac{1}{T\hat{\sigma}^3} \sum_{t=1}^{T} \left(X_t - \bar{X}\right)^3 \\ \widehat{\mathcal{K}}_T &= \frac{1}{T\hat{\sigma}^4} \sum_{t=1}^{T} \left(X_t - \bar{X}\right)^4 \end{aligned} \tag{15}$$

Moment-based tests for normality may be based on both of the above statistics. For example, a test that the skewness of the data is zero is based on the fact that under the null $\sqrt{\frac{T}{6}}\widehat{\mathcal{S}}_T$ follows the standard normal distribution and hence $\frac{T}{6}\widehat{\mathcal{S}}_T^2$ follows a χ_1^2 (chi-squared distribution with one degree of freedom). Similarly, under the null that the kurtosis is equal to three, the statistic $\sqrt{\frac{T}{24}}\left(\widehat{\mathcal{K}}_T - 3\right)$ follows the standard normal distribution and hence $\frac{T}{24}\left(\widehat{\mathcal{K}}_T - 3\right)^2$ follows a χ_1^2 distribution. To increase power, it is natural to consider combining information in the skewness and kurtosis statistics. The Jarque-Bera test (see e.g., Bera and Jarque, 1982) does exactly this, employing the result that the test statistics based on the skewness and kurtosis coefficients are asymptotically independent. The Jarque-Bera test statistic is

$$\mathcal{JB}_T = T\left[\frac{\widehat{\mathcal{S}}_T^2}{6} + \frac{\left(\widehat{\mathcal{K}}_T - 3\right)^2}{24T}\right], \tag{16}$$

which under the null follows a χ^2_2.[20]

1.2 Robust tests for normality

A potential disadvantage of tests based on the skewness or kurtosis coefficient, as well as the Jarque-Bera test, is that such tests are highly sensitive to outliers. Brys, Hubert and Struyf (hereafter "BHS", 2003, 2004a, 2004b and 2005) propose robust measures of skewness and tail weight and suggest a robustified test version of the Jarque-Bera test. The authors propose the *medcouple* ($\mathcal{MC}$) as a robust measure of skewness and the right- and left-medcouples ($\mathcal{RMC}$ and $\mathcal{LMC}$) as robust tail weight measures. Letting F denote the distribution of X_t, $\mathcal{MC}$ is defined as

$$\mathcal{MC}(F) = med_{x_1 \leq m_F \leq x_2} h(x_1, x_2) \tag{17}$$

for x_1 and x_2 sampled from F, $m_F = F^{-1}(0.5)$ and kernel function $h()$ given by

$$h(x, x) = \frac{(x_j - m_F) - (x_i - m_F)}{x_j - x_i}.$$

The $\mathcal{RMC}$ and $\mathcal{LMC}$ are defined by the medcouple on the upper and lower halves of the distribution of X_t, respectively:

$$\mathcal{LMC}(F) = -\mathcal{MC}(x < m_F) \text{ and } \mathcal{RMC}(F) = \mathcal{MC}(x > m_F).$$

Finite sample versions of these quantities, denoted $\mathcal{MC}_T$, $\mathcal{RMC}_T$ and $\mathcal{LMC}_T$ are constructed using the empirical distribution function in place of the population distribution function F. Computation of these quantities may be performed in $O(T \log T)$ time using algorithms described by BHS (2004a). Finally, the quantities depend only on the quantiles of the distribution and thus may be computed for any distribution irrespective of the existence of moments (see, e.g, BHS, 2003).

BHS (2004b) propose various tests for normality based on medcouples. The tests are Wald-type tests based on a set of statistics $w = (\omega_1, .., \omega_k)'$ such that

$$\sqrt{T} w \sim N_k(\omega, \Sigma_k)$$

so that a general test statistic for the normal may be based on

[20] The test statistics reported in the paper are based on skewness and kurtosis statistics that are adjusted for bias and which therefore differ slightly from statistics based on the formula in (15). This distinction does not materially affect the results.

$$T(w-\omega)'\Sigma_k^{-1}(w-\omega) \approx \chi_k^2. \tag{18}$$

BHS (2004b) propose tests of the form (18) based on $w = MC$, $w = (LMC, RMC)'$ and $w = (MC, LMC, RMC)'$, which we refer to as *RJB1 ("Robust Jarque-Bera 1")*, *RJB2* and *RJB3*, respectively.To implement the tests, we must obtain ω and Σ_k under the null that the data follow the normal distribution. These quantities are tabulated in BHS (2004b). Finally, as discussed by BHS (2004b), while these tests are less sensitive to outlying values then the corresponding tests based skewness and kurtosis coefficients, they are less powerful to detect non-normality than the classical tests.

References

Ait-Sahalia, Yacine, 2002, Telling from discrete data whether the underlying continuous-time model is a diffusion, Journal of Finance, 57, 2975-2112.

Ait-Sahalia, Yacine, 2004, Disentangling diffusion from jumps, Journal of Financial Economics, 74, 487-528.

Andersen, Torben G., and Tim Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, Journal of Empirical Finance 4, 115–158.

Andersen, Torben G., and Tim Bollerslev, 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, International Economic Review 39, 885–905.

Andersen, Torben G., Tim Bollerslev and Francis X. Diebold, 2004, Some like it smooth and some like it rough: Untangling continuous and jump components in measuring, modeling and forecasting asset return volatility, Working paper, Wharton School, University of Pennsylvania.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens, 2001, The distribution of realized stock return volatility, Journal of Financial Economics 61, 43–76.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2000, Great realizations, Risk Magazine 13, 105–108.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2001, The distribution of realized exchange rate volatility, Journal of the American Statistical Association 96, 42–55.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and forecasting realized volatility, Econometrica 71, 579–625.

Andreou, Elena, and Eric Ghysels, 2002, Rolling-sample volatility estimators: Some new theoretical, simulation, and empirical results, Journal of Business and Economic Statistics 20, 363–376.

Back, Kerry, 1991, Asset prices for general processes, Journal of Mathematical Economics, 20, 317-395.

Bandi, Federico and Jeffrey Russell, 2004, Separating market microstructure noise from volatility, Journal of Financial Economics, Forthcoming.

Barndorff-Nielsen, Ole E., Peter R. Hansen, Asger Lunde and Neil Shephard, 2004, Regular and modified kernel-based estimators of integrated variance: The case with independent noise, Working paper, Nuffield College, Oxford University.

Barndorff-Nielsen, Ole E., and Neil Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society, Series B 64, 253–280.

Barndorff-Nielsen, Ole E. and Neil Shephard, 2005, Variation, jumps, market frictions and high-frequency data in financial econometrics, Working paper, Nuffield College, Oxford University.

Barndorff-Nielsen, Ole E. and Neil Shephard, 2006, Econometrics of testing for jumps in financial economics using bipower variation, Journal of Financial Econometrics, Forthcoming.

Bollerslev, Tim, and Dan Jubinski, 1999, Equity trading volume and volatility: Latent information arrivals and common long-run dependencies, Journal of Business and Economic Statistics 17, 9–21.

Bollerslev, Tim, Uta Kretschmer, Christian Pigorsch and George Tauchen, 2005, A discrete-time model for daily S&P 500 returns and realized variances: Jumps and leverage effects, Working paper, Duke University.

Bontemps, Christian, and Nour Meddahi, Testing normality: A GMM approach, Journal of Econometrics, Forthcoming.

Brys, G., M. Hubert and A. Struyf, 2003, A comparison of some new measures of skewness. In: Developments in Robust Statistics, ICORS 2001, R. Dutter, P. Filzmoser, U. Gather and P.J. Rousseeuw (eds.), Springer-Verlag, Heidelberg, 98-113.

Brys, G., M. Hubert and A. Struyf, 2004a, A robust measure of skewness. Journal of Computational and Graphical Statistics, 13, 996-1017.

Brys, G., M. Hubert and A. Struyf, 2004b, A robustification of the Jarque-Bera test of normality, COMPSTAT 2004 proceedings, edited by J. Antoch, Springer, Physica Verlag, pp. 753-760, ISBN 3-7908-1554-3.

Brys, G., M. Hubert and A. Struyf, 2005, Robust measures of tail weight, Computational Statistics and Data Analysis, Forthcoming.

Chernov, Michael, A. Ronald Gallant, Eric Ghysels and George Tauchen, 2003, Alternative models for stock price dynamics, Journal of Econometrics, 116, 225-257.

Duffie, Darrell, D. Pan and Kenneth Singleton, 2000, Transform analysis and option pricing for affine-jump diffusions. Econometrica, 68, 1343-1377.

Eraker, Bjorn., Michael Johannes and Nicholas Polson, 2003, The impact of jumps in equity index volatility and returns, Journal of Finance, 2003, 53, p. 1269-1300.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 2004, Stochastic volatility, trading volume, and the daily flow of information, Journal of Business, Forthcoming.

Fleming, Jeff and Bradley S. Paye, 2005, Optimizing Realized Bipower Variation in the Presence of Microstructure Noise, Working notes, Jones Graduate School of Management, Rice University.

Granger, Clive W. J., 2005, The present and future of empirical finance, Financial Analysts Journal, 61, 15-18.

Hansen, Peter R., and Asger Lunde, 2006, Realized Variance and Market Microstructure Noise, Journal of Business and Economic Statistics, The 2005 Invited Address, Forthcoming.

Huang, Xin and George Tauchen, 2005, The relative contribution of jumps to total price variance, Journal of Financial Econometrics, Forthcoming.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.

Oomen, Roel C. A., 2002, Modeling realied variance when returns are serially correlated, Working paper, Warwick Business School, University of Warwick.

Oomen, Roel C. A., 2004, Properties of bias-corrected realized variance in calendar time and business time, Working paper, Warwick Business School, University of Warwick.

Shao, Jun, 1999, Mathematical Statistics, Springer-Verlag, New York, ISBN 0-387-98674-X.

Tauchen, George, Harold Zhang, and Ming Liu, 1996, Volume, volatility, and leverage: A dynamic analysis, Journal of Econometrics 74, 177–208.

Tauchen, George and Hao Zhou, 2005, Identifying realized jumps on financial markets, Working paper, Duke University and Board of Governors of the Federal Reserve.

Zhang, Lan, Per A. Mykland and Yacine Ait-Sahalia, 2004, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, Journal of the American Statistical Association, Forthcoming.

Table 1
Basic summary statistics

The table reports daily summary statistics for the 20 MMI stocks. For each stock, we report the number of trading days in the sample period, the average number of trades per day, the average time (in seconds) to the last transaction price before each 5-minute gridpoint, the average daily open-to-close squared return, and the annualized volatility rate. The sample period is January 4, 1993 to December 31, 2003.

Ticker	Days	Average # of trades	Ave time to 5-min gridpoint	Average daily squared return	Annualized volatility
AXP	2,771	1,726	34	0.000368	30.4%
CVX	2,771	1,389	38	0.000185	21.6%
DD	2,771	1,625	31	0.000284	26.8%
DIS	2,771	2,806	19	0.000321	28.5%
DOW	2,771	965	54	0.000312	28.1%
EK	2,771	1,181	36	0.000269	26.0%
GE	2,771	5,533	13	0.000260	25.6%
GM	2,771	1,738	23	0.000311	28.0%
IBM	2,771	3,740	13	0.000341	29.3%
IP	2,771	1,138	58	0.000328	28.8%
JNJ	2,771	2,291	20	0.000209	23.0%
KO	2,771	2,265	18	0.000217	23.4%
MCD	2,771	1,826	24	0.000254	25.3%
MMM	2,771	1,196	46	0.000198	22.4%
MO	2,770	3,011	15	0.000324	28.6%
MRK	2,771	2,966	13	0.000258	25.5%
PG	2,771	1,862	27	0.000221	23.6%
S	2,771	1,165	39	0.000436	33.2%
T	2,771	3,425	13	0.000368	30.5%
XOM	2,771	2,467	22	0.000180	21.3%

Table 2
Serial correlation of returns at different sampling frequencies

The table reports the average number of returns per day and the serial correlation of returns for different sampling frequencies. We compute the intradaily returns for each of the 20 MMI stocks using sampling frequencies ranging from 30 seconds to 60 minutes. For each sampling frequency, we compute the first- through fifth-order serial correlation coefficients for each stock, and we report the average coefficients in the table. The sample period is January 4, 1993 to December 31, 2003 (2,771 trading days for most stocks).

Frequency (in minutes)	Average # of returns	Serial correlation coefficients				
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
0.5	772	−0.148	−0.036	−0.012	−0.005	−0.002
1.0	386	−0.143	−0.013	−0.003	−0.004	−0.003
1.5	257	−0.123	−0.007	−0.006	−0.004	−0.003
2.0	193	−0.105	−0.009	−0.006	−0.005	−0.003
2.5	155	−0.092	−0.010	−0.006	−0.005	−0.003
3.0	129	−0.084	−0.010	−0.006	−0.004	−0.001
3.5	111	−0.077	−0.011	−0.005	−0.003	0.001
4.0	97	−0.073	−0.010	−0.006	0.000	0.002
4.5	86	−0.069	−0.011	−0.004	0.001	0.001
5.0	77	−0.069	−0.010	−0.001	0.002	0.001
6.0	65	−0.061	−0.009	0.001	0.002	0.004
7.0	56	−0.057	−0.006	0.003	0.003	0.003
8.0	49	−0.054	−0.002	0.003	0.005	0.003
9.0	44	−0.050	−0.002	0.006	0.003	0.005
10.0	39	−0.048	0.000	0.008	0.004	0.004
15.0	26	−0.035	0.008	0.007	0.005	0.002
20.0	20	−0.023	0.011	0.007	0.003	0.004
30.0	13	−0.005	0.012	0.006	0.007	0.001
45.0	9	0.003	0.010	0.008	−0.003	0.007
60.0	7	0.012	0.008	0.003	0.006	−0.012

Table 3
Summary statistics for realized variance

The table reports summary statistics for realized variance and log realized variance for the 20 MMI stocks. The realized variances are constructed using a five-minute sampling frequency. For each stock, we report the mean, standard deviation, skewness, and kurtosis of the daily realized variance and the daily log realized variance. The sample period is January 4, 1993 to December 31, 2003 (2,771 trading days for most stocks).

	Realized variance				Log realized variance			
Ticker	Mean	Std	Skew	Kurt	Mean	Std	Skew	Kurt
AXP	0.00046	0.00043	4.75	39.57	−7.93	0.70	−0.06	4.37
CVX	0.00020	0.00014	3.00	19.73	−8.69	0.62	−0.05	3.31
DD	0.00033	0.00028	3.45	22.71	−8.25	0.66	0.34	3.35
DIS	0.00042	0.00044	8.39	147.34	−8.03	0.67	0.44	3.74
DOW	0.00031	0.00043	9.81	189.20	−8.50	0.83	0.60	3.39
EK	0.00030	0.00032	5.97	58.33	−8.36	0.65	0.78	4.40
GE	0.00030	0.00036	9.12	165.81	−8.42	0.72	0.52	3.81
GM	0.00031	0.00025	4.35	34.30	−8.30	0.61	0.27	4.09
IBM	0.00032	0.00032	5.81	74.26	−8.31	0.71	0.28	3.47
IP	0.00039	0.00034	3.13	19.69	−8.14	0.73	0.15	3.01
JNJ	0.00026	0.00022	5.96	72.54	−8.44	0.62	0.10	3.99
KO	0.00028	0.00021	3.88	28.75	−8.38	0.60	0.22	3.82
MCD	0.00035	0.00030	5.24	50.40	−8.14	0.59	0.44	3.95
MMM	0.00023	0.00022	3.65	24.82	−8.64	0.73	0.14	3.38
MO	0.00034	0.00049	12.00	235.62	−8.33	0.74	0.56	4.16
MRK	0.00030	0.00026	5.13	50.48	−8.31	0.62	0.19	3.93
PG	0.00027	0.00026	5.35	53.62	−8.49	0.69	0.16	3.98
S	0.00042	0.00040	4.51	36.36	−8.02	0.66	0.58	3.71
T	0.00038	0.00041	5.13	45.62	−8.17	0.71	0.68	3.42
XOM	0.00022	0.00022	5.92	65.18	−8.65	0.64	0.65	3.91

Table 4

Returns standardized by realized volatility

The table reports summary statistics and the results of normality tests for returns standardized by realized volatility. The returns are daily open-to-close returns and the realized volatilities are constructed using a five-minute sampling frequency. For each stock, we report the mean, standard deviation, skewness, and kurtosis of the standardized returns, and the p-values for normality tests based on the Jarque-Bera (JB), JB skewness (JB-skew), JB kurtosis (JB-kurt), as well as the robust versions of the JB tests (RbstJB). The sample period is January 4, 1993 to December 31, 2003 (2,771 trading days for most stocks).

					Test statistics (p-values)					
Ticker	Mean	Std	Skew	Kurt	JB	JB-skew	JB-kurt	RbstJB1	RbstJB2	RbstJB3
AXP	0.01	0.89	0.12	3.10	0.01	0.01	0.17	0.86	0.70	0.84
CVX	−0.02	0.94	0.07	2.97	0.24	0.09	0.81	0.45	0.51	0.66
DD	0.02	0.89	0.18	2.94	0.01	0.01	0.13	0.05	0.09	0.02
DIS	−0.02	0.86	0.09	2.84	0.10	0.05	0.40	0.88	0.52	0.73
DOW	0.03	0.95	0.05	3.03	0.54	0.28	0.83	0.10	0.01	0.02
EK	−0.04	0.91	0.14	3.08	0.03	0.02	0.24	0.01	0.72	0.04
GE	0.02	0.92	0.16	2.83	0.01	0.00	0.38	0.05	0.17	0.03
GM	−0.06	1.00	0.16	2.84	0.01	0.00	0.19	0.00	0.82	0.00
IBM	−0.01	1.00	0.13	2.65	0.00	0.02	0.00	0.54	0.11	0.12
IP	−0.05	0.91	0.18	2.98	0.01	0.00	0.50	0.00	0.83	0.02
JNJ	0.02	0.87	0.06	2.87	0.41	0.22	0.60	0.97	0.92	0.98
KO	0.05	0.86	0.13	3.18	0.00	0.00	0.00	0.06	0.76	0.22
MCD	0.03	0.82	0.08	3.12	0.11	0.05	0.43	0.18	0.34	0.16
MMM	−0.01	0.90	0.06	3.04	0.50	0.85	0.24	0.86	0.74	0.88
MO	0.03	0.91	0.12	2.93	0.05	0.01	0.79	0.47	0.51	0.64
MRK	0.04	0.92	0.12	2.87	0.03	0.01	0.65	0.03	0.76	0.10
PG	0.09	0.87	0.08	2.84	0.02	0.03	0.07	0.90	0.45	0.65
S	0.01	0.96	0.02	2.78	0.04	0.70	0.01	0.88	0.00	0.01
T	−0.06	0.93	0.14	3.10	0.00	0.00	0.20	0.03	0.29	0.03
XOM	0.02	0.87	0.04	2.88	0.71	0.46	0.71	0.49	0.16	0.28

Table 5

Returns standardized by Newey-West realized volatility

The table reports summary statistics and the results of normality tests for returns standardized by a Newey-West estimator of realized volatility. The returns are daily open-to-close returns and the realized volatilities are constructed using a 30-second sampling frequency and a 30-minute window width. For each stock, we report the mean, standard deviation, skewness, and kurtosis of the standardized returns, and the p-values for normality tests based on the Jarque-Bera (JB), JB skewness (JB-skew), JB kurtosis (JB-kurt), as well as the robust versions of the JB tests (RbstJB). The sample period is January 4, 1993 to December 31, 2003 (2,771 trading days for most stocks).

					Test statistics (p-values)					
Ticker	Mean	Std	Skew	Kurt	JB	JB-skew	JB-kurt	RbstJB1	RbstJB2	RbstJB3
AXP	0.00	0.94	0.04	2.33	0.00	0.40	0.00	0.93	0.00	0.00
CVX	−0.02	0.96	0.02	2.36	0.00	0.59	0.00	0.42	0.01	0.02
DD	0.02	0.95	0.13	2.37	0.00	0.01	0.00	0.09	0.00	0.00
DIS	−0.04	0.94	0.05	2.32	0.00	0.27	0.00	0.49	0.01	0.02
DOW	0.02	0.96	0.00	2.43	0.00	0.94	0.00	0.13	0.01	0.01
EK	−0.05	0.93	0.07	2.48	0.00	0.11	0.00	0.01	0.00	0.00
GE	0.01	0.97	0.09	2.31	0.00	0.06	0.00	0.64	0.00	0.00
GM	−0.07	1.00	0.14	2.32	0.00	0.00	0.00	0.00	0.01	0.00
IBM	−0.02	1.02	0.09	2.26	0.00	0.05	0.00	0.74	0.00	0.00
IP	−0.07	0.95	0.16	2.41	0.00	0.00	0.00	0.00	0.02	0.00
JNJ	0.01	0.94	0.02	2.38	0.00	0.73	0.00	0.76	0.00	0.00
KO	0.05	0.93	0.05	2.41	0.00	0.29	0.00	0.05	0.05	0.02
MCD	0.02	0.91	0.02	2.45	0.00	0.74	0.00	0.44	0.01	0.01
MMM	−0.03	0.92	0.03	2.45	0.00	0.54	0.00	0.97	0.03	0.07
MO	0.02	0.97	0.05	2.27	0.00	0.26	0.00	0.29	0.04	0.06
MRK	0.04	0.98	0.06	2.31	0.00	0.22	0.00	0.08	0.00	0.00
PG	0.09	0.94	-0.02	2.37	0.00	0.71	0.00	0.86	0.00	0.01
S	0.00	0.99	0.01	2.27	0.00	0.90	0.00	0.69	0.00	0.01
T	−0.08	0.99	0.13	2.32	0.00	0.00	0.00	0.81	0.00	0.00
XOM	0.02	0.96	0.03	2.32	0.00	0.49	0.00	0.70	0.03	0.06

Table 6
Summary statistics for large-return days

The table reports summary statistics for the 20 MMI stocks on days with absolute open-to-close returns greater than 2 percent. For each stock, we report the number of high-return days, and the average number of trades, the average time (in seconds) to the last transaction price before each 5-minute gridpoint, the average daily open-to-close squared return, and the annualized volatility rate on the large-return days. The sample period is January 4, 1993 to December 31, 2003.

Ticker	Days	Average # of trades	Ave time to 5-min gridpoint	Average daily squared return	Annualized volatility
AXP	668	2,214	22	0.001199	55.0%
CVX	366	1,640	28	0.000846	46.2%
DD	523	2,150	19	0.001093	52.5%
DIS	609	3,923	12	0.001109	52.9%
DOW	507	1,354	34	0.001320	57.7%
EK	484	1,678	25	0.001110	52.9%
GE	496	9,101	6	0.001037	51.1%
GM	617	2,105	19	0.001038	51.1%
IBM	647	4,734	9	0.001134	53.5%
IP	615	1,391	35	0.001127	53.3%
JNJ	417	2,740	16	0.000875	47.0%
KO	394	3,142	12	0.000983	49.8%
MCD	480	2,535	16	0.001043	51.3%
MMM	353	1,500	30	0.000975	49.6%
MO	536	4,392	10	0.001293	57.1%
MRK	517	3,684	10	0.000958	49.1%
PG	393	2,616	20	0.001007	50.4%
S	763	1,514	30	0.001307	57.4%
T	615	5,108	8	0.001321	57.7%
XOM	316	3,421	12	0.000921	48.2%

Table 7
Serial correlation of intraday returns on large-return days

The table reports the average number of returns per day and the serial correlation of returns for different sampling frequencies on large-return days. We compute the intradaily returns for each of the 20 MMI stocks using sampling frequencies ranging from 30 seconds to 60 minutes. For each sampling frequency, we compute the first- through fifth-order serial correlation coefficients for each stock, and we report the average coefficients in the table. The sample period is January 4, 1993 to December 31, 2003.

Frequency (in minutes)	Average # of returns	Serial correlation coefficients				
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
0.5	773	−0.134	0.005	0.010	0.010	0.006
1.0	387	−0.072	0.019	0.011	0.004	0.001
1.5	258	−0.037	0.019	0.004	0.004	0.004
2.0	194	−0.015	0.011	0.006	0.004	0.004
2.5	155	−0.005	0.008	0.007	0.003	0.006
3.0	129	0.001	0.008	0.006	0.007	0.011
3.5	111	0.004	0.008	0.009	0.008	0.013
4.0	97	0.007	0.009	0.008	0.015	0.015
4.5	86	0.008	0.008	0.013	0.018	0.016
5.0	78	0.005	0.011	0.019	0.018	0.017
6.0	65	0.012	0.013	0.023	0.023	0.024
7.0	56	0.014	0.017	0.030	0.025	0.024
8.0	49	0.015	0.027	0.030	0.030	0.024
9.0	44	0.019	0.031	0.037	0.027	0.029
10.0	39	0.024	0.033	0.039	0.030	0.030
15.0	26	0.040	0.052	0.046	0.040	0.034
20.0	20	0.063	0.065	0.055	0.044	0.042
30.0	13	0.090	0.078	0.065	0.064	0.058
45.0	9	0.120	0.096	0.087	0.082	0.112
60.0	7	0.153	0.110	0.113	0.147	0.179

Table 8

Returns standardized by realized bipower variation

The table reports summary statistics and the results of normality tests for returns standardized by realized bipower variation. The returns are daily open-to-close returns and the realized bipower variation is constructed using a 10-minute sampling frequency. For each stock, we report the mean, standard deviation, skewness, and kurtosis of the standardized returns, and the p-values for normality tests based on the Jarque-Bera (JB), JB skewness (JB-skew), JB kurtosis (JB-kurt), as well as the robust versions of the JB tests (RbstJB). The sample period is January 4, 1993 to December 31, 2003 (2,771 trading days for most stocks).

					Test statistics (p-values)					
Ticker	Mean	Std	Skew	Kurt	JB	JB-skew	JB-kurt	RbstJB1	RbstJB2	RbstJB3
AXP	0.01	0.98	0.08	2.92	0.12	0.06	0.40	0.83	0.48	0.65
CVX	−0.02	1.04	0.06	2.87	0.09	0.24	0.06	0.64	0.61	0.78
DD	0.03	0.98	0.12	2.89	0.10	0.18	0.09	0.05	0.00	0.00
DIS	−0.03	0.96	0.07	2.81	0.09	0.04	0.46	0.85	0.87	0.96
DOW	0.03	1.04	0.06	2.96	0.28	0.16	0.44	0.10	0.21	0.20
EK	−0.04	1.00	0.10	3.03	0.07	0.02	0.91	0.01	0.96	0.05
GE	0.02	1.02	0.17	2.89	0.00	0.00	0.64	0.06	0.02	0.00
GM	−0.06	1.09	0.15	2.69	0.00	0.01	0.01	0.00	0.21	0.00
IBM	−0.02	1.10	0.12	2.71	0.01	0.04	0.03	0.64	0.08	0.11
IP	−0.07	1.01	0.13	3.06	0.07	0.03	0.42	0.01	0.13	0.01
JNJ	0.02	0.98	0.07	2.83	0.13	0.12	0.19	1.00	0.04	0.09
KO	0.06	0.97	0.15	2.99	0.00	0.00	0.60	0.05	0.98	0.22
MCD	0.03	0.93	0.07	3.02	0.12	0.04	0.68	0.17	0.90	0.58
MMM	−0.02	0.98	0.05	2.93	0.87	0.78	0.65	1.00	0.35	0.55
MO	0.03	1.01	0.11	2.79	0.02	0.02	0.13	0.50	0.09	0.16
MRK	0.04	1.03	0.10	2.74	0.01	0.02	0.06	0.04	0.46	0.17
PG	0.10	0.97	0.08	2.84	0.02	0.03	0.09	0.66	0.07	0.11
S	0.01	1.07	0.02	2.68	0.01	0.68	0.00	0.81	0.00	0.01
T	−0.07	1.04	0.17	2.84	0.00	0.00	0.43	0.22	0.37	0.20
XOM	0.02	0.99	0.03	2.78	0.20	0.81	0.08	0.50	0.28	0.35

Table 9

Returns standardized by realized bipower variation: Pre- vs. post-decimalization

The table reports summary statistics and the results of normality tests for returns standardized by realized bipower variation for the pre- and post-decimalization sample periods. The returns are daily open-to-close returns and the realized bipower variation is constructed using a 10-minute sampling frequency. For each stock, we report the standard deviation, skewness, and kurtosis of the standardized returns, and the p-values for normality tests based on the Jarque-Bera (JB), JB skewness (JB-skew), and JB kurtosis (JB-kurt) test statistics. The pre-decimalization sample period is March 4, 1998 to January 26, 2001, and the post-decimalization sample period is January 29, 2001 to December 31, 2003 (733 trading days in each sample period).

	Pre-decimalization						Post-decimalization					
Ticker	Std	Skew	Kurt	JB	JB-skew	JB-kurt	Std	Skew	Kurt	JB	JB-skew	JB-kurt
AXP	1.10	0.17	2.53	0.01	0.06	0.01	1.04	0.03	2.88	0.76	0.70	0.52
CVX	1.10	0.10	2.70	0.14	0.26	0.10	1.02	0.04	3.08	0.81	0.64	0.64
DD	1.03	0.07	3.03	0.74	0.45	0.89	0.98	0.26	3.12	0.01	0.00	0.51
DIS	1.00	0.14	2.76	0.14	0.14	0.19	0.98	−0.03	2.55	0.04	0.75	0.01
DOW	1.12	0.15	2.72	0.07	0.09	0.12	1.02	0.03	3.09	0.85	0.75	0.64
EK	0.96	0.20	2.86	0.07	0.03	0.45	0.96	0.01	3.04	0.96	0.89	0.82
GE	1.07	0.12	3.04	0.43	0.20	0.81	1.06	0.19	2.75	0.04	0.03	0.17
GM	1.15	0.07	2.63	0.09	0.43	0.04	1.08	0.18	2.59	0.01	0.05	0.02
IBM	1.06	0.15	2.78	0.13	0.10	0.22	1.09	0.13	2.56	0.02	0.16	0.02
IP	1.01	0.16	2.60	0.02	0.08	0.03	1.00	0.16	2.97	0.21	0.08	0.86
JNJ	1.06	0.14	2.77	0.14	0.13	0.20	0.94	0.03	2.70	0.25	0.76	0.10
KO	1.04	0.19	2.82	0.07	0.04	0.32	0.96	0.05	2.81	0.51	0.62	0.29
MCD	1.02	0.03	2.58	0.07	0.73	0.02	0.94	−0.04	3.21	0.45	0.64	0.24
MMM	0.98	−0.10	2.78	0.24	0.25	0.22	1.02	0.11	2.84	0.30	0.21	0.36
MO	1.09	0.08	2.67	0.14	0.40	0.07	1.03	0.07	2.67	0.15	0.45	0.07
MRK	1.10	0.12	2.59	0.04	0.20	0.02	1.02	−0.01	2.53	0.03	0.88	0.01
PG	1.04	0.02	2.80	0.54	0.86	0.27	0.93	−0.02	2.90	0.83	0.81	0.58
S	1.11	0.07	2.63	0.10	0.44	0.04	1.04	0.03	2.74	0.36	0.78	0.16
T	1.09	0.19	2.52	0.00	0.03	0.01	1.08	−0.03	2.70	0.24	0.70	0.10
XOM	1.01	0.09	2.49	0.01	0.34	0.01	0.97	0.08	2.71	0.19	0.39	0.11

Table 10

Non-jump component of returns standardized by realized bipower variation

The table reports summary statistics and the results of normality tests for the non-jump component of returns standardized by realized bipower variation. We estimate the jump component on day n as

$$J_n = sign(r_n) \times \sqrt{(RV_n - BV_n) \times I_n\{\text{test rejects}\}}$$

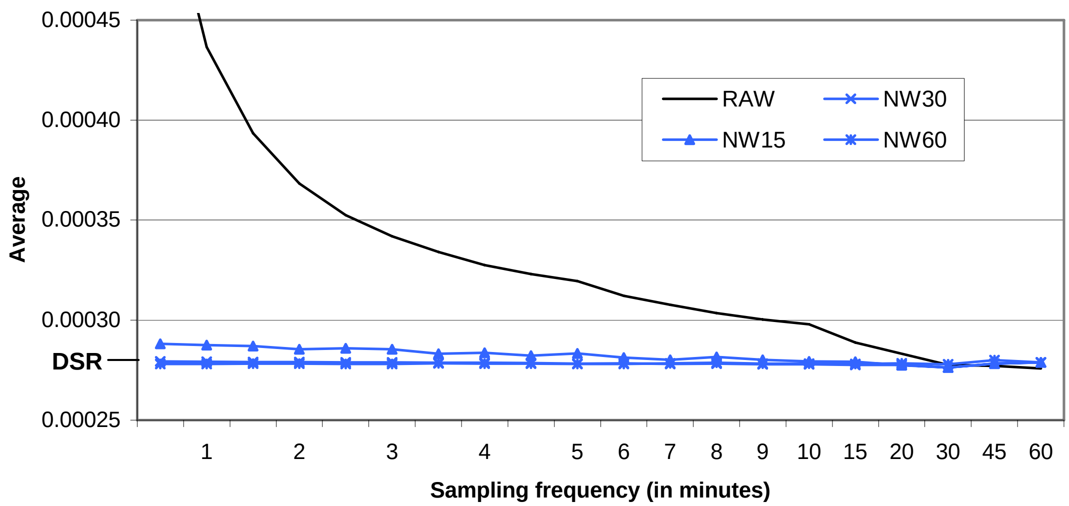
where $I_n\{\text{test rejects}\}$ is a one on days when the null of no jumps is rejected at the 1% level and zero otherwise. We then construct the non-jump component of returns by subtracting J_n from the daily open-to-close return, and we construct the realized bipower variation using a 10-minute sampling frequency. For each stock, we report the mean, standard deviation, skewness, and kurtosis of the standardized returns, and the p-values for normality tests based on the Jarque-Bera (JB), JB skewness (JB-skew), JB kurtosis (JB-kurt), as well as the robust versions of the JB tests (RbstJB). The sample period is January 4, 1993 to December 31, 2003 (2,771 trading days for most stocks).

					Test statistics (p-values)					
Ticker	Mean	Std	Skew	Kurt	JB	JB-skew	JB-kurt	RbstJB1	RbstJB2	RbstJB3
AXP	0.00	0.85	0.05	2.61	0.00	0.15	0.00	0.86	0.31	0.45
CVX	−0.03	0.86	0.04	2.71	0.01	0.29	0.00	0.26	0.82	0.67
DD	0.01	0.85	0.19	2.74	0.00	0.00	0.00	0.12	0.48	0.25
DIS	−0.03	0.84	0.08	2.68	0.01	0.14	0.00	0.14	0.92	0.47
DOW	0.01	0.85	0.01	2.80	0.08	0.59	0.03	0.54	0.72	0.76
EK	−0.04	0.83	0.03	2.78	0.07	0.67	0.02	0.56	0.22	0.35
GE	0.02	0.89	0.10	2.63	0.00	0.09	0.00	0.27	0.22	0.17
GM	−0.06	0.89	0.13	2.59	0.00	0.00	0.00	0.00	0.95	0.00
IBM	−0.02	0.91	0.12	2.55	0.00	0.01	0.00	0.09	0.36	0.18
IP	−0.06	0.85	0.18	2.70	0.00	0.00	0.01	0.00	0.56	0.00
JNJ	0.01	0.84	0.03	2.72	0.05	0.61	0.02	0.70	0.90	0.93
KO	0.04	0.83	0.10	2.74	0.00	0.03	0.01	0.06	0.87	0.25
MCD	0.02	0.81	0.05	2.80	0.04	0.28	0.02	0.31	0.97	0.76
MMM	−0.03	0.83	0.03	2.75	0.14	0.91	0.05	0.41	0.67	0.75
MO	0.01	0.87	0.05	2.54	0.00	0.27	0.00	0.91	0.03	0.08
MRK	0.03	0.87	0.04	2.61	0.00	0.20	0.00	0.09	0.58	0.27
PG	0.08	0.84	0.04	2.64	0.00	0.29	0.00	0.09	0.26	0.12
S	−0.01	0.89	0.01	2.61	0.00	0.84	0.00	0.37	0.00	0.00
T	−0.07	0.89	0.14	2.61	0.00	0.00	0.00	0.24	0.16	0.11
XOM	0.01	0.86	0.05	2.62	0.00	0.44	0.00	0.44	0.70	0.73

Figure 1
Mean and standard deviation of realized variance

The figure plots the mean and standard deviation of the daily realized variance as a function of sampling frequency. We construct the daily realized variance for each of the 20 MMI stocks during the sample period January 4, 1993 to December 31, 2003 (2,771 observations for most stocks). We construct both RAW estimates and NW estimates based on window widths of 15, 30 and 60 minutes (i.e., NW15, NW30 and NW60), using sampling frequencies ranging from 30 seconds to 60 minutes. For each set of realized variances, we compute the mean and standard deviation, and the figure plots the average mean (Panel A) and standard deviation (Panel B) for the 20 stocks. The tick mark labeled "DSR" on the y-axis in Panel A shows the average daily squared return.

Panel A. Mean realized variance



Panel B. Standard deviation of realized variance

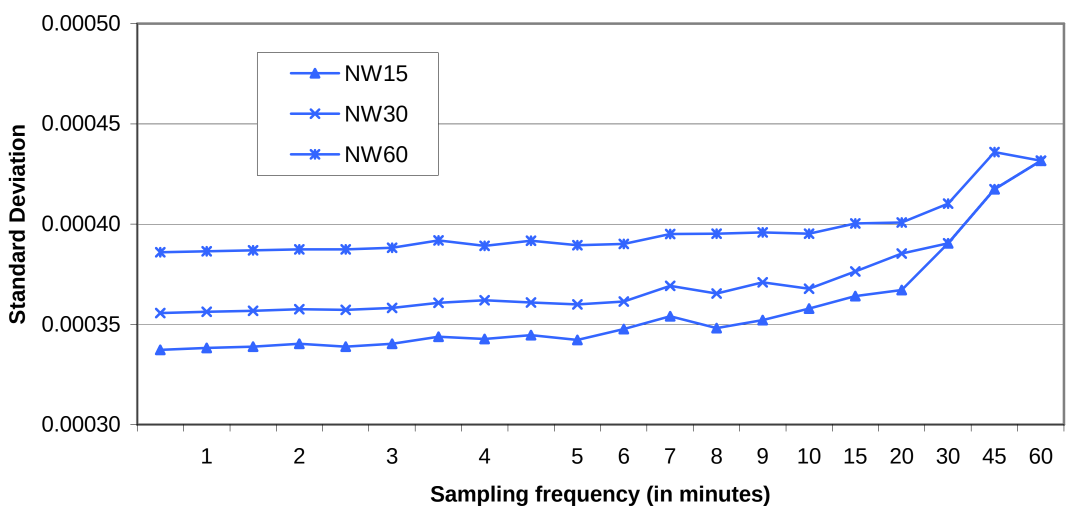
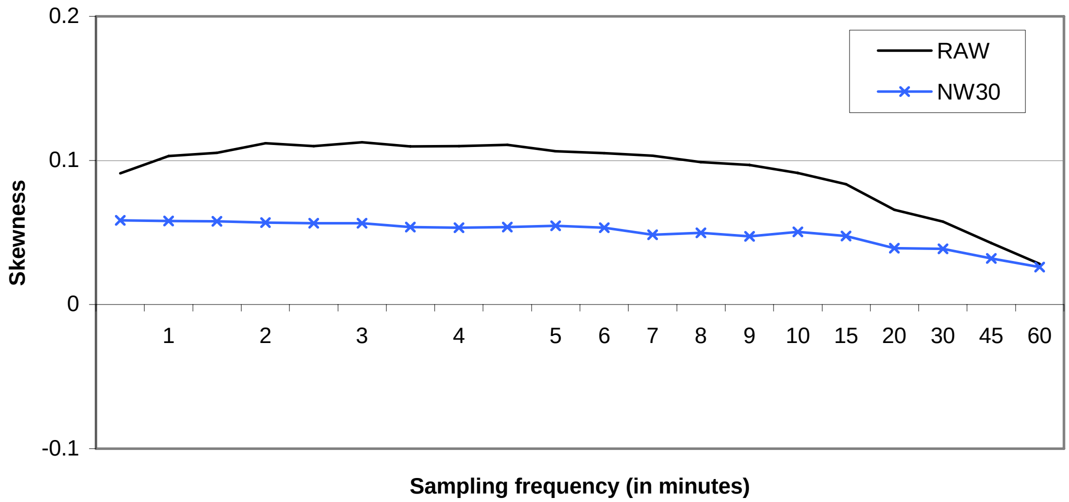


Figure 2
Skewness and kurtosis of returns standardized by realized variance

The figure plots the skewness and kurtosis of daily returns standardized by realized volatility as a function of sampling frequency. We compute the standardized returns for each of the 20 MMI stocks during the sample period January 4, 1993 to December 31, 2003 (2,771 observations for most stocks), using both the RAW and NW30 realized variances constructed using sampling frequencies ranging from 30 seconds to 60 minutes. For each set of returns, we compute the skewness and kurtosis, and the figure plots the average skewness (Panel A) and kurtosis (Panel B) for the 20 stocks.

Panel A. Skewness



Panel B. Kurtosis

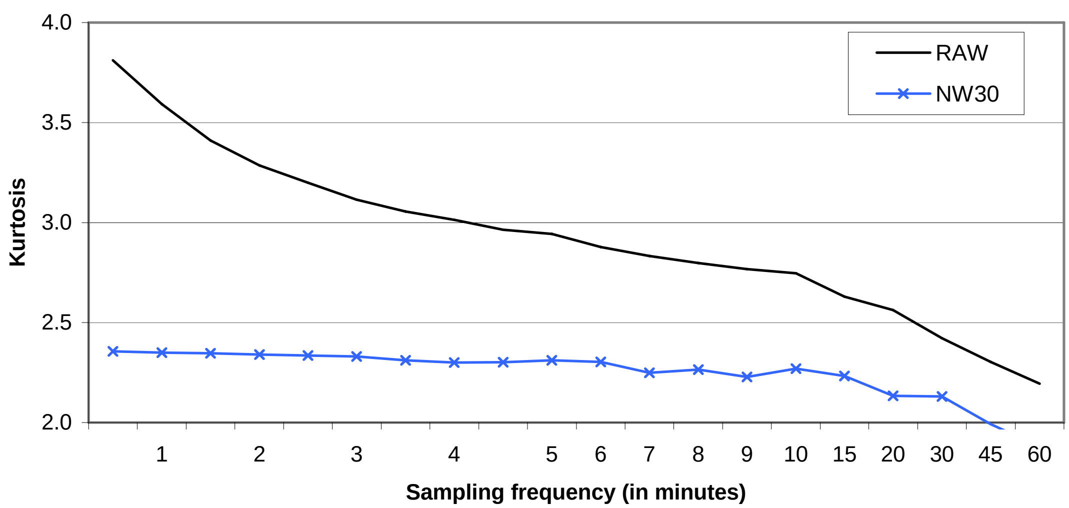


Figure 3
Mean realized variance on large-return days

The figure plots the mean daily realized variance as a function of sampling frequency on large-return days. We construct the daily realized variance for each of the 20 MMI stocks during the sample period January 4, 1993 to December 31, 2003 for those days in which the stock's absolute open-to-close return is greater than 2 percent. The figure shows the mean RAW and NW30 realized variances constructed using sampling frequencies ranging from 30 seconds to 60 minutes.

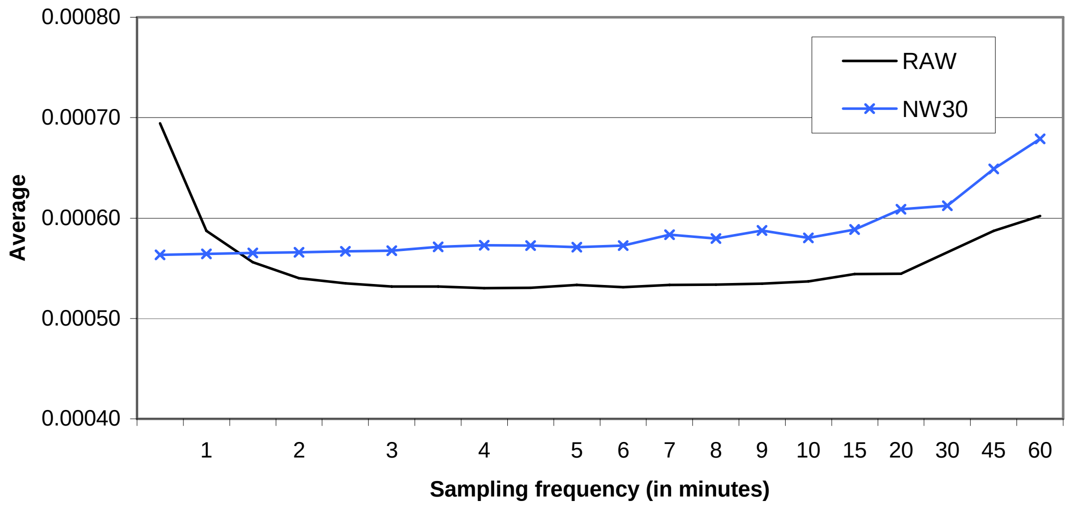


Figure 4
Realized bipower variation signature plot

The figure plots realized bipower variation as a function of sampling frequency. We construct the daily realized bipower variation (BV) for each of the 20 MMI stocks during the sample period January 4, 1993 to December 31, 2003 (2,771 observations for most stocks) using sampling frequencies from 30 seconds to 60 minutes. For each sampling frequency, we compute the mean realized bipower variation for the 20 stocks. For comparison, we also report the mean (RAW) realized variance. The thick line in the figure denotes the mean realized bipower variation using half-day returns.

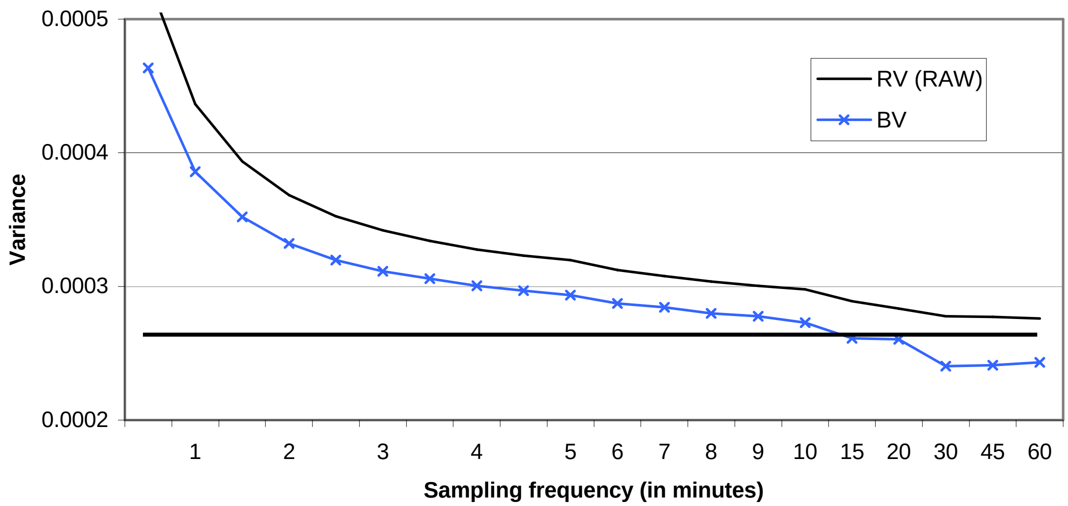
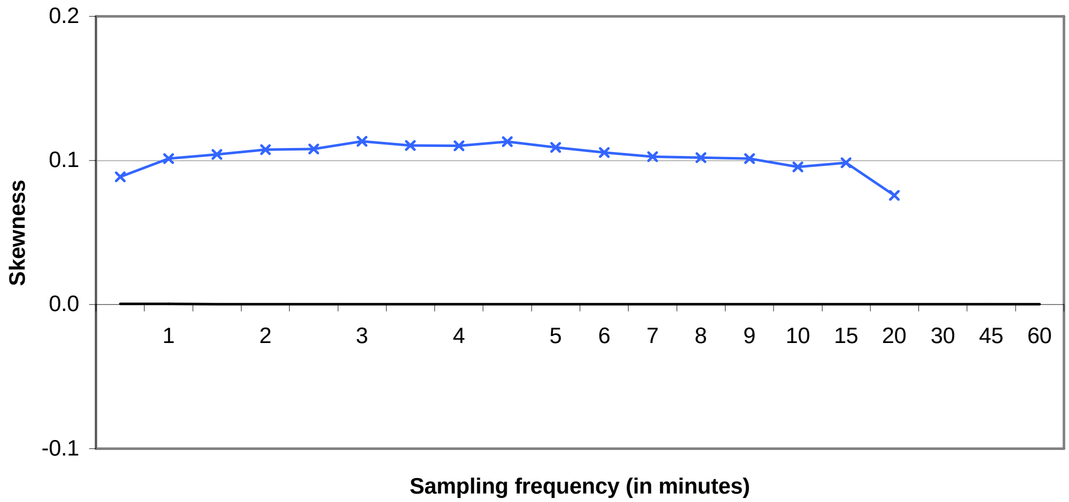


Figure 5

Skewness and kurtosis of returns standardized by realized bipower variation

The figure plots the skewness and kurtosis of daily returns standardized by realized bipower variation as a function of sampling frequency. We compute the standardized returns for each of the 20 MMI stocks during the sample period January 4, 1993 to December 31, 2003 (2,771 observations for most stocks), using realized bipower variation constructed using sampling frequencies ranging from 30 seconds to 60 minutes. For each set of returns, we compute the skewness and kurtosis, and the figure plots the average skewness (Panel A) and kurtosis (Panel B) for the 20 stocks.

Panel A. Skewness



Panel B. Kurtosis