Trading activity, realized volatility and jumps

Pierre Giot* Sébastien Laurent† Mikael Petitjean‡

Abstract

This paper takes a new look at the relation between volume and realized volatility. In contrast to prior studies, we decompose realized volatility into two major components: a continuously varying component and a discontinuous jump component. Our results confirm that the number of trades is the dominant factor shaping the volume-volatility relation, whatever the volatility component considered. However, we also show that the decomposition of realized volatility bears on the volume-volatility relation. Trade variables are positively related to the continuous component only. The well-documented positive volume-volatility relation does not hold for jumps.

Keywords: volume, volatility, transactions, jumps, bi-power variation.

JEL classification: G10, G12, G13.

*Louvain School of Management, CORE (Université Catholique de Louvain) and CeReFiM (FUNDP, University of Namur).

†CORE (Université Catholique de Louvain) and CeReFiM (FUNDP, University of Namur).

‡Louvain School of Management and FUCaM (Catholic University of Mons). Send all correspondence to mikael.petitjean@fucam.ac.be.

1 Introduction

Volume and volatility convey extremely important implications for market participants. Not surprisingly, the relation between the two has been extensively studied in the past. In the early empirical literature, which is mostly based on monthly and weekly stock returns, volatility and trading volume are measured respectively by absolute returns and number of shares traded per equally time-spaced intervals. A positive contemporaneous relation between the two is generally documented, although it does not always appear to be sizeable. Evidence can be found in Karpoff (1987), Jain and Joh (1988), Schwert (1989), Lamoureux and Lastrapes (1990), and Gallant, Rossi, and Tauchen (1992).

In more recent theoretical studies, volume is decomposed into trade frequency (i.e. the number of trades) and trade size (i.e. the average number of shares per trade). A rough taxonomy of the theoretical models on the volume-volatility relation consists of three classes of models: competitive, strategic, and mixture of distributions models.

Competitive microstructure models are extensions of the Glosten and Milgrom (1985) sequential trading model in which market makers and uninformed investors experience adverse selection when trading with informed investors. In this model, each investor is not allowed to transact more than one unit of stock per unit of time, hence (absolute) price changes are independent of trade size. In contrast, Easley and O'Hara (1987) allow traders to transact varying trade sizes and allow for uncertainty in the information arrival process of the informed trader. When investors differ in their beliefs regarding the importance of information and act competitively, larger-sized trades tend to be executed by better-informed investors: larger trades exhibit a greater adverse selection effect. Thus, competitive models suggest a positive relation between trade size and price volatility (Easley, Kiefer, and O'Hara, 1997; Shalen, 1993; Harris and Raviv, 1993; Kim and Verrecchia, 1991; Holthausen and Verrecchia, 1990; Grundy and McNichols, 1989).

Strategic microstructure models also incorporate asymmetric information across agents but assume that informed investors engage in stealth trading by breaking up large trades into many smaller transactions. Therefore, the effect of trade size on price volatility is attenuated and its impact may be transferred to the number of trades. Besides, market makers tend to infer the information content of a trade from order imbalance because they cannot distinguish whether a specific order comes from an informed or uninformed trader. Thus, strategic models predict a positive relation between volatility and number of trades and/or order imbalance (Holden and Subrahmanyam, 1992; Foster and Viswanathan, 1990; Admati and Pfleiferer, 1988; Kyle, 1985).

The third class of models relies on the mixture of distributions (MD) hypothesis (Harris, 1987; Tauchen and Pitts, 1983; Epps and Epps, 1976; Clark, 1973). Although MD models have been criticized on the grounds that they are primarily statistical models (as opposed to economical equilibrium models), recent empirical works have provided strong evidence coherent with the predictions of MD models. For example, these models predict that average trade sizes should have no effect on price volatility. In contrast, it is the number of trades, rather than the total volume, that should reflect the number of daily information arrivals. Supportive evidence is given by Andersen (1996) who tests the MD model based on the microstructure framework of Glosten and Milgrom (1985), and by Ane and Geman (2000) who use high frequency data on actively traded stocks.

Our own research builds on four empirical papers. First, Jones, Kaul, and Lipson (1994) decompose trading volume into its two components and find that stock price volatility is driven by the number of trades per equally time-spaced intervals. The average trade size offers no additional explanatory power beyond the information conveyed by the number of trades. Second, Chan and Fong (2000) filter the effects of order imbalance on returns and find that number of trades explains very little of the absolute residuals. They conclude that it is order imbalance, rather than number of trades, that drives the volume-volatility relation. Third, Huang and Masulis (2003) distinguish between small and large trades. For large trades, they confirm the Jones, Kaul, and Lipson (1994) findings, that only trade frequency affects price volatility. For small trades, the picture is roughly similar with the exception of small trades close to the maximum-guaranteed quoted depth, for which trade frequency *and* average trade size impact price volatility. Fourth, in contrast to the three above-mentioned papers, Chan and Fong (2006) use realized volatility in place of absolute returns as the volatility measure. Absolute returns is indeed a very noisy estimator of the true latent volatility. Since daily absolute returns are computed using only two prices (opening and closing), the computed volatility may be very low if the opening and closing price are very close, even though there might be significant intraday price fluctuations. Chan and Fong (2006) find that neither trade size nor order imbalance adds significantly more explanatory power to realized volatility beyond number of trades. Unfortunately, Chan and Fong (2006) use transaction prices to estimate realized volatility. However, transaction prices are much more affected by residual noise than bid-ask midquotes since transaction prices suffer from bid-ask bounce effects while midquotes do not (Hansen and Lunde, 2006; Bandi and Russell, 2006).

This paper innovates in several ways. First, our sample covers a five-year period and includes the 100 largest stocks quoted on the New York Stock Exchange as of January 1,

1995. This is the most extensive data set used to study the relationship between volume and volatility. Second, we use midquotes to measure prices, in agreement with the recent literature on realized variance. Third, and most importantly, our study is the first to decompose realized volatility into its continuous and jump components. Since the identification of actual jumps is not readily available from the time-series data of underlying asset returns, the empirical estimation of the jump-diffusion processes has always been a challenge in finance. Most of the econometric work relies on some combination of numerical methods, computationally intensive simulation-based procedures, and possibly joint identification schemes from both the underlying asset and the derivative prices. This paper takes a different and direct approach to identify the realized jumps, based on the Barndorff-Nielsen and Shephard (2004, 2006) bi-power variation method. Bi-power variation delivers consistent estimates of continuous volatility, even in the presence of jumps. In addition, realized volatility (i.e. the sum of squared intradaily returns) is a consistent estimate of the sum of both continuous volatility and jumps in the underlying price process. Therefore, the difference between realized volatility and bi-power variation consistently estimates the contribution of discontinuities (i.e. jumps) to the quadratic variation process.

The fact that realized volatility can be decomposed prompts the following question: would the results of prior studies still hold if one decomposes realized volatility into its continuous, persistent part and its discontinuous, temporary, jump component? The motivation behind the decomposition of realized volatility into its two components relies on the following observation: the positive volatility-volume relation that is documented in the literature focuses exclusively on the *level* of volatility (i.e. low versus high volatility). The *nature* of volatility (i.e. continuous versus discontinuous volatility) is completely ignored, and one way to characterize the nature of volatility is precisely to introduce the concept of jumps. This is especially important since market participants usually care as much about the nature of volatility as about its level. For example, all traders make the distinction between 'good' and 'bad' volatilities. 'Good' volatility is directional, persistent, relatively easy to anticipate and accompanied by sufficiently high volume. 'Bad' volatility is *jumpy*, relatively difficult to foresee and associated with low volume. As such, 'good' and 'bad' volatilities can be respectively associated with the continuous, persistent part and the discontinuous, jump component of realized volatility.

Our findings can be summarized as follows. Based on 5 years of intraday data, we show that the decomposition of realized volatility bears on the volume-volatility relation. Trade variables are positively related to the continuous, persistent component only. The positive relationship between volume and volatility does not hold for jumps, supporting the view of market participants according to which poor trading volume leads to more erratic price changes. This negative volume-jumps relation is revealed through the number of trades, which remains the dominant factor behind the volume-volatility relation. Beyond the number of trades, neither trade size nor order imbalance increases explanatory power significantly, whatever the volatility component considered.

The remainder of the paper is organized as follows. In Section 2, we explain the procedure used to identify the jump component of realized volatility. In Section 3, we describe the data and provide summary statistics. In Section 4, we report and interpret the empirical results. We conclude in Section 5.

2 The decomposition of realized volatility

To decompose realized volatility into its continuous and jump components, we consider the following continuous-time jump diffusion process:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \qquad 0 \leq t \leq, T \tag{1}$$

where $p(t)$ is a logarithmic asset price at time t, $\mu(t)$ is a continuous and locally bounded variation process, $\sigma(t)$ is a strictly positive stochastic volatility process with a sample path that is right continuous and has well defined limits, $W(t)$ is a standard Brownian motion, and $q(t)$ is a counting process with intensity $\lambda(t)$ ($P[dq(t) = 1] = \lambda(t)dt$ and $\kappa(t) = p(t) - p(t-)$ is the size of the jump in question). The quadratic variation for the cumulative process $r(t) \equiv p(t) - p(0)$ is the integrated volatility of the continuous sample path component plus the sum of the $q(t)$ squared jumps that occurred between time 0 and time t:

$$[r, r]_t = \int_0^t \sigma^2(s)ds + \sum_{0<s\leq t} \kappa^2(s). \tag{2}$$

Now, let us define the daily realized volatility as the sum of the corresponding intradaily squared returns:

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r^2_{t+j\Delta,\Delta}, \tag{3}$$

where $r_{t,\Delta} \equiv p(t) - p(t - \Delta)$ is the discretely sampled Δ-period return.[1] So $1/\Delta$ is the number of intradaily periods.

Barndorff-Nielsen and Shephard (2004) show that the realized volatility converges uni-

[1] We use the same notation as in Andersen, Bollerslev, and Diebold (2006) and normalize the daily time interval to unity. We drop the Δ subscript for daily returns: $r_{t+1,1} \equiv r_{t+1}$.

formly in probability to the increment of the quadratic variation process as the sampling frequency of the returns increases ($\Delta \rightarrow 0$):[2]

$$RV_{t+1}(\Delta) \rightarrow \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s\leq t+1} \kappa^2(s). \tag{4}$$

That implies that the realized volatility is a consistent estimate for integrated volatility as long as there are no jumps.

In order to disentangle the continuous and the jump components of realized volatility, we need to consistently estimate integrated volatility, even in the presence of jumps in the process. This is done using the asymptotic results of Barndorff-Nielsen and Shephard (2004, 2006). The realized bi-power variation is defined as the sum of the product of adjacent absolute intradaily returns standardized by a constant:

$$BV_{t+1}(\Delta) \equiv \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta,\Delta}||r_{t+(j-1)\Delta,\Delta}|, \tag{5}$$

where $\mu_1 \equiv \sqrt{2/\pi} \simeq 0.79788$ is the mean of the absolute value of a standard normally distributed random variable. It can indeed be shown that even in the presence of jumps,

$$BV_{t+1}(\Delta) \rightarrow \int_t^{t+1} \sigma^2(s)ds. \tag{6}$$

Thus, the difference between the realized volatility and the bi-power variation consistently

[2] See also, for example, Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b), Comte and Renault (1998).

estimates the jump contribution to the quadratic variation process. When $\Delta \to 0$:

$$RV_{t+1}(\Delta) - BV_{t+1}(\Delta) \to \sum_{t<s\leq t+1} \kappa^2(s). \tag{7}$$

Moreover, because a finite sample estimate of the squared jump process might be negative (in Equation 7), we truncate the measurement at zero, i.e.

$$J_{t+1}(\Delta) \equiv max[RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0]. \tag{8}$$

One might wish to select only statistically significant jumps, i.e. to consider very small jumps as being part of the continuous sample path rather than genuine discontinuities. The Barndorff-Nielsen and Shephard (2004, 2006) results, extended in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2005), imply:

$$\frac{RV_{t+1}(\Delta) - BV_{t+1}(\Delta)}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2} - 5)\Delta \int_t^{t+1} \sigma^4(s)ds}} \to N(0,1), \tag{9}$$

when there is no jump and for $\Delta \to 0$, under sufficient regularity conditions. We need to estimate the integrated quarticity $\int_t^{t+1} \sigma^4(s)ds$ to compute this statistic. The realized tri-power quarticity measure permits us to estimate it consistently, even in the presence of jumps:

$$TQ_{t+1}(\Delta) \equiv \Delta^{-1}\mu_{4/3}^{-3} \sum_{j=3}^{1/\Delta} |r_{t+j\Delta,\Delta}|^{4/3}|r_{t+(j-1)\Delta,\Delta}|^{4/3}|r_{t+(j-2)\Delta,\Delta}|^{4/3}, \tag{10}$$

with $\mu_{4/3} \equiv 2^{2/3}\Gamma(7/6)\Gamma(1/2)^{-1}$. Thus, we have that, for $\Delta \to 0$:

$$TQ_{t+1}(\Delta) \rightarrow \int_t^{t+1} \sigma^4(s)ds. \tag{11}$$

The implementable statistics is therefore:

$$W_{t+1}(\Delta) \equiv \frac{RV_{t+1}(\Delta) - BV_{t+1}(\Delta)}{\sqrt{\Delta(\mu_1^{-4} + 2\mu_1^{-2} - 5)TQ_{t+1}(\Delta)}}. \tag{12}$$

However, following Huang and Tauchen (2005) and Andersen, Bollerslev, and Diebold (2006), we actually compute the following statistic:

$$Z_{t+1}(\Delta) \equiv \Delta^{-1/2} \frac{[RV_{t+1}(\Delta)) - BV_{t+1}(\Delta)]RV_{t+1}(\Delta)^{-1}}{[(\mu_1^{-4} + 2\mu_1^{-2} - 5)max\{1, TQ_{t+1}(\Delta)BV_{t+1}(\Delta)^{-2}\}]^{1/2}}. \tag{13}$$

Huang and Tauchen (2005) show that the statistic defined in Equation (12) tends to over-reject the null hypothesis of no jumps. Moreover, they show that $Z_{t+1}(\Delta)$ defined in Equation (13) is closely approximated by a standard normal distribution and has reasonable power against several plausible stochastic volatility jump diffusion models. Practically, we choose a significance level α and compute:

$$J_{t+1,\alpha}(\Delta) = I[Z_{t+1}(\Delta) > \Phi_\alpha] \cdot [RV_{t+1}(\Delta) - BV_{t+1}(\Delta)]. \tag{14}$$

Of course, a smaller α means that we estimated fewer and larger jumps. Moreover, to ensure that the sum of the jump and continuous components equals the realized volatility, we impose:

$$C_{t+1,\alpha}(\Delta) = I[Z_{t+1}(\Delta) \leq \Phi_\alpha] \cdot RV_{t+1}(\Delta) + I[Z_{t+1}(\Delta) > \Phi_\alpha] \cdot BV_{t+1}(\Delta). \tag{15}$$

Finally, as suggested by Andersen, Bollerslev, and Diebold (2006), we use 'staggered' versions of the bi-power variation and tri-power quarticity measure to tackle microstructure noise that causes the high-frequency returns to be autocorrelated. Basically, these 'staggered' versions amount to skip one observation when computing the product of adjacent returns. If the order of the serial correlation is higher than one, one may choose to skip more than one return. The staggered versions of bi-power variation and tri-power quarticity are respectively given by:

$$BV_{t+1}(\Delta) \equiv \mu_1^{-2}(1-2\Delta)^{-1} \sum_{j=3}^{1/\Delta} |r_{t+j\Delta,\Delta}||r_{t+(j-2)\Delta,\Delta}|, \tag{16}$$

$$TQ_{t+1}(\Delta) \equiv \Delta^{-1}\mu_{4/3}^{-3}(1-4\Delta)^{-1} \sum_{j=5}^{1/\Delta} |r_{t+j\Delta,\Delta}|^{4/3}|r_{t+(j-2)\Delta,\Delta}|^{4/3}|r_{t+(j-4)\Delta,\Delta}|^{4/3}. \tag{17}$$

3 Data

The sample period covers a five-year period starting on January 1, 1995 and ending on September 30, 1999, which represents a total of 1199 trading days. The sample consists of the 100 largest stocks traded on the New York Stock Exchange as of January 1, 1995. We choose the 100 largest stocks because these stocks are actively traded, yield sufficiently high frequency intraday returns for computing reliable daily realized volatility estimates, and are likely to benefit from high information arrival rates.

Data for this study is retrieved from the Trades and Quotes (TAQ) database. Following Chordia, Roll, and Subrahmanyam (2000), we only retain class A stocks and remove preferred stocks or shares, warrants, rights, derivatives, trusts, closed-end investment companies, American depositary receipts, units, shares of beneficial interest, holdings and realty trusts. We restrict our selection to stocks whose price is higher than \$5 and lower than \$999. The remaining stocks are then selected on the basis of market capitalization. Applying traditional filtering procedures (Chordia, Roll, and Subrahmanyam, 2001; Huang and Stoll, 1996), we reject quotes exhibiting (a) price (at the bid or at the ask) lower than or equal to 0; (b) size (at the bid or at the ask) lower than or equal to 0; (c) price at the bid higher than price at the ask; (d) bid-ask spread greater than \$4; (e) proportional bid-ask spread greater than 40%. Trades are excluded if they satisfy at least one of the following conditions: (a) trade price is lower than or equal to 0; (b) trade size is lower than or equal to 0; (c) trade is not "regular", i.e. it is subsequently corrected or canceled. We additionally remove any trade or quote time-stamped outside regular trading hours, that is, before 9:30 AM and after 4:00 PM (or 1:00 PM on the days the exchange closed early). We also exclude the opening transaction for each day. Finally, following Chordia, Roll, and Subrahmanyam (2001), we exclude records for which the (proportional) effective spread was greater than four times the (proportional) quoted spread.

In agreement with the extant literature, which utilizes high-frequency return data for the purpose of volatility estimation, we use mid-quote prices to construct returns and then to compute realized volatility estimators, as described in Equation (3). Finally, realized volatility is decomposed into the continuous sample path and the jump component as described respectively in equations (14) and (15). The significance of the jump component is assessed using a conservative 99.99 % confidence level (i.e. $\alpha = 0.9999$).

Following the early literature, we decompose daily trading volume into number of trades and average trade size (ratio of number of shares traded to number of trades). Following Chan and Fong (2006), we also include order imbalance. We compute daily absolute order imbalance as the absolute value of the number of buyer-initiated trades minus number of seller initiated trades for the day. To define the direction of a trade, we rely upon the widely-used Lee and Ready (1991) algorithm.[3] We follow the recommendation contained in *SEC Rule 11Ac1-5*, assuming trades were recorded 5 seconds later than their actual execution time.

Table 1 gives the cross-sectional means of time series statistics for the 7 variables used in our study. Using the square root of time, the daily realized mean variance (RV) of 0.0262% translates into an annualized realized volatility of 25.6%. This is close to the annualized realized volatility of 27.5% reported by Chan and Fong (2006) for the Dow 30 stocks, over the 1993-2000 period. We also find very similar autocorrelation coefficients. We nevertheless report a higher coefficient of variation (97%) than the one reported by Chan and Fong (34%). This can be explained by the stronger heterogeneity of stocks in our sample. Following Beine, Lahaye, Laurent, Neely, and Palm (2007), we use a conservative significance level (i.e. $\alpha = 0.9999$) to identify the number of economically meaningful jumps (given by $J9999$ in Table 1). The proportion of days with significant jumps is 26%, i.e. 313 days over 1199 on average. RV and the daily realized continuous variance display similar characteristics. Interestingly, all variables, except jumps, have similar autocorrelation patterns. Among the trade variables, average trade size (ATS)

[3]Odders-White (2000) and Ellis, Michaely, and O'Hara (2000) document that the Lee and Ready algorithm correctly classifies more than 80% of the transactions in their respective samples and that trades executed inside the spread are more likely to be misclassified. Lee and Radhakrishna (2000) also show that classification errors arising from the Lee and Ready algorithm are small enough not to present serious problems. They find that the algorithm accurately classifies 92% of the trades classifiable in the TORQ database, with 96% accuracy on trades executed at the bid or ask price, 70% accuracy on trades executed inside the spread but not at the quote midpoint, and 65% accuracy on trades executed at the quote midpoint.

displays the lowest AR and CV coefficients. Daily number of trades (NT) exhibits the slowest decaying autocorrelation function while absolute order imbalance $|OB|$ displays the highest CV.

Table 2 reports correlations between volatility and transactions. We find very similar results to those reported by Chan and Fong (2006). More precisely, we confirm that: trading volume is highly correlated with number of trades; number of trades co-varies quite strongly with order imbalance; realized volatility correlates more highly with number of trades than with average trade size or order imbalance. Interestingly, the jump component ($J9999$) does not behave like the undecomposed measure of realized volatility (RV). First, the two measures are negatively correlated with one another. Second, volume (V) and number of trades (NT) are negatively correlated with $J9999$, while they are positively correlated with RV. This suggests that the decomposition of realized volatility may bear on the volume-volatility relation. As pointed out in previous studies, the overall level of realized volatility increases (decreases) when trading volume (V) and, in particular, number of trades (NT) increase (decrease). However, the decomposition of realized volatility seems to indicate that such a positive volume-volatility relation does not hold when the jump component of realized volatility is considered. It would thus only hold for the continuous part.

4 Empirical Analysis

In the following analysis, we estimate the volume-volatility relation for each of our 100 stocks. Following Chan and Fong (2006) (CF), we estimate this relation by OLS, use auto-correlation and heteroskedasticity-robust (Newey and West) standard errors, and include a

Monday dummy as well as 12 lags to account for some dynamics in the conditional expected return. Since Huang and Masulis (2003) (HM) advocate the use of Hansen's (1982) generalized method of moments (GMM) method, we also estimate the relationship by GMM, which imposes weak distribution assumptions on the observable variables and endogenously adjusts the estimates to account for general forms of conditional heteroskedasticity and/or serial correlation that may be present in the error structure. We follow the CF and HM approaches to regress realized volatility (RV) and its continuous component ($C9999$) on various trade measures, i.e. number of shares traded (V), number of trades (NT), average number of shares per trade (ATS), and absolute order imbalance ($|OB|$). However, we do not follow these two approaches to estimate the relationship between jumps and trading activity. We instead run Tobit regressions since jumps have values censored at zero. The population distribution of jumps is indeed spread over a large range of positive values, with a pileup at the value zero. Tables 3 to 8 report the impact of the above trade measures on realized volatility and its two components.[4]

Table 3 presents the results for trading volume (i.e. the number of shares traded). When the CF approach is followed with realized volatility as the dependent variable (i.e. Regression 1), results are qualitatively similar to those reported by Chan and Fong (2006). The R^2 is slightly higher, up from around 36% to 40%. The percentage of stocks for which volume is statistically significant rises from 95% to 100%. The mean ($\hat{\phi}$) is equal to 0.11, implying that an increase in the number of shares traded of 100,000 is accompanied by an increase in realized volatility of around 5%. The p-value is particularly low, equal to 0.05%. Results do not fundamentally change when the HM regression (i.e. Regression (3),

[4]To test whether our results were induced by the upward trend in trading activity, we added a time trend in each of the following regressions. Since results were almost identical, we do not report them to save space. They are available upon request. We also applied Phillips-Perron unit root tests to test for stochastic trends in the de-trended series. There was no evidence of stochastic trends.

estimated by GMM) is used: while the R^2 is lower, 99 stocks still display positive and significant coefficients for trading volume. Results are very similar when the continuous component of realized volatility is used as the dependent variable in Regressions (2) and (4). For jumps, however, the picture is completely different. Regression (5), i.e. the Tobit regression, indicates that jumps and trading volume are significantly and negatively related for 72 stocks out of 100. Only 2 stocks displays positive and significant coefficients for trading volume at the 5% level.

Table 4 reports the results for trade frequency (i.e. the number of trades). Like trading volume, number of trades explains a substantial portion of daily realized volatility. In the 'CF regression', i.e. Regression (1), the mean ($\hat{\beta}$) is equal to 0.318, implying that an increase in the number of trades of 100 is accompanied by an increase in realized volatility of around 9%. We obtain similar results by using the 'HM regression', i.e. Regression (3). Overall, the use of number of trades instead of number of shares traded as explicative variable leads to a rise in the R^2, which is more pronounced in the HM regression. Again, number of trades affects jumps differently. Regression (5) shows that jumps and number of trades are significantly and negatively related for 86 stocks out of 100. Only 3 stocks displays positive and significant coefficients for trading volume at the 5% level. The average p-value is around 3%, down from around 8% in the specification with trading volume. Trade frequency appears to be more informative than trading volume in explaining both jumps and the continuous component of realized volatility.

The effect of average trade size (i.e. the average number of shares per trade) is shown in Table 5. In line with previous studies, trade size explains much less realized volatility than trading volume or trade frequency. The explanatory power of trade size is poor: no single p-value is lower than 20%. Overall, the use of average trade size as a regressor leads to a fall in the R^2, which is more severe in the HM regression estimated by GMM. In addition, the HM regression indicates that the percentage of stocks with positive and significant coefficients for trade size is equivalent to the percentage of stocks displaying *negative* and significant coefficients. For jumps, the average ($\hat{\gamma}$) coefficient is positive but not significant, confirming the poor explanatory power of trade size.

Table 6 shows that adding average trade size to number of trades has no significant effect on the adjusted R^2. When compared to Table 4, the adjusted R^2 slightly increases in the CF regressions, while it slightly decreases in the HM regressions. Moreover, adding average trade size in the regression does not impact the explanatory power of number of trades. All coefficients for number of trades remain significant. The poor explanatory power of average trade size is also confirmed. As in Table 5, the average p-value does not even reach the 20% level. In the HM specification for both realized volatility and its continuous component, only a quarter of trade size coefficients are positive as well significantly different from zero at the 5% level, confirming the results of Table 5. The results are even worst when jumps are considered. Average trade size clearly plays no role in explaining jumps and number of trades remains the dominant factor.

Table 7 reports the results for absolute order imbalance as sole explicative variable. There are at least two reasons why order imbalances can provide additional power beyond pure trading activity measures (such as trade frequency or/and trade size). First, a high absolute order imbalance can alter returns as market makers struggle to re-adjust their inventory. Second, order imbalances can signal excessive investor interest in a stock, and if this interest is autocorrelated, then order imbalances could be related to future returns.

Indeed, Chordia and Subrahmanyam (2004) state that "Intuition suggests that the implications of a reported volume of one million shares generated by 500,000 shares of seller initiated trades and 500,000 shares of buyer-initiated trades are very different from those generated by one million shares of seller- (or buyer-) initiated trades." Table 7 shows that absolute order imbalance does have a role in explaining realized volatility. However, evidence is less convincing for order imbalance than for number of trades. First, the lowest average p-value is still above 5%. Second, the number of stocks displaying positive and significant coefficients for order imbalance is around 20% lower than the number of stocks displaying positive and significant coefficients for number of trades (see Table 4). Third, and most importantly, the explanatory power of order imbalance is much weaker once realized volatility is decomposed. This is particularly obvious for jumps. While number of trades was significant at the 5% level in the jump equation (see Regression (5) of Table 4), this is not the case for order imbalance: The average ($\hat{\lambda}$) coefficient in Regression (5) is not significant, even at 10%. Finally, the percentage of stocks with negative and significant coefficients for order imbalance is 41 only: it was twice higher for number of trades.

Table 8 reports the results when absolute order imbalance and number of trades are both included as regressors. Adding absolute order imbalance to number of trades does not help. First, while the adjusted R^2's for the CF regressions are similar to those reported in Table 4, they decrease when order imbalance is added to number of trades in the HM regressions. Second, no average p-value for order imbalance is lower than 25%. Third, only one single stock exhibits a positive and significant coefficient for order imbalance when the dependent variable is either realized volatility or its continuous component. The results are similar for jumps: number of trades dominates order imbalance. Number of trades is, on average, significant at 5% while order imbalance is not. Number of trades is also the most pervasive factor, as jumps and number of trades are significantly and negatively related for 88 stocks out of 100.

There is strong evidence that trade frequency remains the dominant factor behind the volume-volatility relation. The decomposition of realized volatility even strengthens the dominant role played by number of trades. However, trade frequency affects the two components of realized volatility very differently: the relation between trade frequency and the continuous component of realized volatility is positive while the relation between trade frequency and the jump component is negative.

5 Conclusion

The ability to identify realized jumps has important implications in financial management, from portfolio and risk management to option and bond pricing and hedging (Merton, 1976; Bates, 1996; Bakshi, Cao, and Chen, 2000; Liu, Longstaff, and Pan, 2003; Duffie and Singleton, 2000; and Piazzesi, 2003). In particular, Eraker, Johannes, and Polson (2003) show that jump components command relatively larger risk premia than continuous components, as their contribution to periods of market stress is greater. There is therefore a practical interest in identifying jumps.

By decomposing realized volatility into its two components, we show that trading activity relates to volatility in a more subtle way than previously thought. Because volatility can be diffusive or discontinuous, stating that 'the greater the level of volume, the greater the volatility' hides a more complex reality. While trading activity relates positively to diffusive (or continuous) volatility, it relates negatively to jumps (or discontinuous volatility).

In other words, the positive relationship between volume and volatility that is documented in the literature only holds for diffusive volatility. This indicates that poor trading volume leads to more erratic volatility changes, as commonly argued in dealing rooms. This negative volume-jumps relation is revealed through the number of trades, which remains the dominant factor behind the volume-volatility relation. Neither trade size nor order imbalance adds significantly more explanatory power beyond number of trades, whatever the volatility component considered.

Further empirical work on jumps needs to be done. While Andersen, Bollerslev, Diebold, and Vega (2007) show that jumps in exchange rates, stocks and bonds are linked to fundamentals, we still need to understand how the link really works. This will certainly help revive studies about the impact of news and economic events on financial markets.

References

Admati, A., and P. Pfleiferer (1988): "A theory of intraday patterns: volume and price variability," *Review of Financial Studies*, 1, 3–40.

Andersen, T. (1996): "Return volatility and trading volume: an information flow interpretation of stochastic volatility," *Journal of Finance*, 51, 169–204.

Andersen, T., and T. Bollerslev (1998): "Answering the skeptics: yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39, 885–905.

Andersen, T., T. Bollerslev, F. Diebold, and C. Vega (2007): "Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets," *Journal of International Economics*, 73, 251–277.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2006): "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Ane, T., and H. Geman (2000): "Order flow, transaction clock, and normality of asset returns," *Journal of Finance*, 55, 2259–2284.

Bakshi, G., C. Cao, and Z. Chen (2000): "Pricing and Hedging Long-Term Options," *Journal of Econometrics*, 94, 277–318.

BANDI, F., AND J. RUSSELL (2006): “Separating microstructure noise from volatility,” *Journal of Financial Economics*, 79, 655–692.

BARNDORFF-NIELSEN, O., J. GRAVERSEN, J. JACOD, M. PODOLSKIJ, AND N. SHEPHARD (2005): “A central Limit Theorem for Realized Power and Bipower Variation of Continuous Semimartingale,” in *From Stochastic Analysis to Mathematival Finance, Festschrift for Albert Shiryaev*, ed. by Y. Kabanov, and R. Lipster. Springer Verlag.

BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2002a): “Econometric Analysis of Realised Volatility and its use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society*, 64, 253–280.

——— (2002b): “Estimating Quadratic Variation Using Realized Variance,” *Journal of Applied Econometrics*, 17, 457–478.

——— (2004): “Power and bipower variation with stochastic volatility and jumps (with discussion),” *Journal of Financial Econometrics*, 2, 1–48.

——— (2006): “Econometrics of testing for jumps in financial economics using bipower variation,” *Journal of Financial Econometrics*, 4, 1–30.

BATES, D. S. (1996): “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsch Mark Options,” *Review of Financial Studies*, 9, 69–107.

BEINE, M., J. LAHAYE, S. LAURENT, C. NEELY, AND F. PALM (2007): “Central Bank Intervention and Exchange Rate Volatility, Its Continuous and Jump Components,” *International Journal of Finance and Economics*, 12, 201–223.

Chan, K., and W. Fong (2000): "Trade size, order imbalance, and the volatility-volume relation," *Journal of Financial Economics*, 57, 247–273.

——— (2006): "Realized volatility and transactions," *Journal of Banking and Finance*, 30, 2063–2085.

Chordia, T., R. Roll, and A. Subrahmanyam (2000): "Commonality in liquidity," *Journal of Financial Economics*, 56, 3–28.

——— (2001): "Market liquidity and trading activity," *Journal of Finance*, 56, 501–530.

Chordia, T., and A. Subrahmanyam (2004): "Order imbalance and individual stock returns: Theory and evidence," *Journal of Financial Economics*, 72, 485–518.

Clark, P. (1973): "A subordinated stochastic process model with nite variance for speculative prices," *Econometrica*, 41, 135–155.

Comte, F., and E. Renault (1998): "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323.

Duffie, D. Pan, J., and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions," *Econometrica*, 68, 1343–1376.

Easley, D., N. Kiefer, and M. O'Hara (1997): "One day in the life of a very common stock," *Review of Financial Studies*, 10, 805–835.

Easley, D., and M. O'Hara (1987): "Price, trade size and information in securities markets," *Journal of Financial Economics*, 19, 69–90.

Ellis, K., R. Michaely, and M. O'Hara (2000): "The accuracy of trade classification rules: evidence from Nasdaq," *Journal of Financial and Quantitative Analysis*, 35, 529–551.

Epps, T., and M. Epps (1976): "The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis," *Econometrica*, 44, 305–321.

Eraker, B., M. Johannes, and N. Polson (2003): "The impact of jumps in volatility and returns," *The Journal of Finance*, 58, 1269–1300.

Foster, F., and S. Viswanathan (1990): "A theory of intraday variations in volumes, variances and trading costs in securities markets," *Review of Financial Studies*, 3, 593–624.

Gallant, R., P. Rossi, and G. Tauchen (1992): "Stock prices and volume," *Review of Financial Studies*, 5, 199–242.

Glosten, L., and P. Milgrom (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14, 71–100.

Grundy, B., and M. McNichols (1989): "Trade and the revelation of information through prices and direct disclosure," *Review of Financial Studies*, 2, 495–526.

Hansen, L. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hansen, P., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–218.

Harris, L. (1987): "Transaction data tests of the mixture of distributions hypothesis," *Journal of Financial and Quantitative Analysis*, 22, 127–141.

Harris, M., and A. Raviv (1993): "Differences of opinion make a horse race," *Review of Financial Studies*, 6, 473506.

Holden, C., and A. Subrahmanyam (1992): "Long-lived private information and imperfect competition," *Journal of Finance*, 47, 247–270.

Holthausen, R., and R. Verrecchia (1990): "The effect of informedness and consensus on price and volume behavior," *Accounting Review*, 65, 191–208.

Huang, R., and R. Masulis (2003): "Trading activity and stock price volatility: evidence from the London Stock Exchange," *Journal of Empirical Finance*, 10, 249–269.

Huang, R., and H. Stoll (1996): "Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE," *Journal of Financial Economics*, 41, 313–357.

Huang, X., and G. Tauchen (2005): "The relative contribution of jumps to total price variation," *Journal of Financial Econometrics*, 3, 456–499.

Jain, P., and G. Joh (1988): "The dependence between hourly prices and trading volume," *Journal of Financial and Quantitative Analysis*, 23, 269–284.

Jones, C., G. Kaul, and M. Lipson (1994): "Transactions, volume and volatility," *Rewiew of Financial Studies*, 7, 631–651.

Karpoff, J. (1987): "The relationship between price changes and trading volume: a survey," *Journal of Financial and Quantitative Analysis*, 22, 109–126.

Kim, O., and R. Verrecchia (1991): "Market reaction to anticipated announcements," *Journal of Financial Economics*, 30, 273–309.

Kyle, A. (1985): "Continuous auctions and insider trading," *Econometrica*, 53, 1315–1335.

Lamoureux, C., and W. Lastrapes (1990): "Heteroskedasticity in stock return data: volume versus GARCH effects," *Journal of Finance*, 45, 220–229.

Lee, C., and B. Radhakrishna (2000): "Inferring Investor Behavior: Evidence from TORQ Data," *Journal of Financial Markets*, 3, 183–204.

Lee, C., and M. Ready (1991): "Inferring trade direction from intraday data," *Journal of Finance*, 46, 733–747.

Liu, J., F. Longstaff, and J. Pan (2003): "Dynamic Asset Allocation with Event Risk," *Journal of Finance*, 58, 231–259.

Merton, R. C. (1976): "Option Pricing when Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3, 125–144.

Newey, W., and K. West (1987): "A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Odders-White, E. (2000): "On the occurrence and consequences of inaccurate trade classification," *Journal of Financial Markets*, 3, 259–286.

Piazzesi, M. (2003): "Bond Yields and the Federal Reserve," *Journal of Political Economy*, 113, 311–344.

Schwert, W. (1989): "Why does stock volatility change over time," *Journal of Finance*, 44, 1115–1154.

Shalen, C. T. (1993): "Volume, volatility, and the dispersion of beliefs," *Review of Financial Studies*, 6, 405–434.

Tauchen, G., and M. Pitts (1983): "The price variability-volume relationship on speculative markets," *Econometrica*, 51, 484–505.

Table 1: Summary statistics.

	RV	$J9999$	$C9999$	V	ATS	NT	$\|OB\|$
Mean	2.42	1.04	2.38	1.40	1595	929	121
Standard Deviation	2.56	1.16	2.48	1.00	643	578	153
Skewness	5.13	5.51	4.51	3.40	1.32	2.24	3.67
Kurtosis	56.80	54.69	43.75	27.63	49.67	13.34	31.64
Coefficient of Variation	105.79	111.97	104.20	71.90	39.58	56.32	110.43
AR(1)	0.49	0.06	0.51	0.61	0.39	0.79	0.45
AR(12)	0.23	0.04	0.25	0.31	0.19	0.53	0.21

Realized variance (RV), continuous variance ($C9999$), and significant jumps ($J9999$) are multiplied by 10,000. V is trading volume in millions of shares traded per day, NT denotes daily number of trades, ATS is average trade size (numbers of shares traded each day divided by number of trades for the day), and $|OB|$ is absolute order imbalance defined as the absolute value of the number of buyer-initiated trades minus number of seller initiated trades for the day.

Table 2: Correlation matrix.

	RV	$J9999$	$C9999$	V	ATS	NT	$\|OB\|$
RV	1.00						
$J9999$	-0.22	1.00					
$C9999$	0.92	-0.53	1.00				
V	0.54	-0.21	0.54	1.00			
ATS	0.05	0.02	0.03	0.49	1.00		
NT	0.62	-0.27	0.63	0.80	-0.04	1.00	
$\|OB\|$	0.22	0.07	0.21	0.34	-0.05	0.42	1.00

RV =realized variance; $J9999$ = significant jumps at $\alpha = 0.9999$; $C9999$ = continuous variance with $\alpha = 0.9999$; V = trading volume; NT = daily number of trades; ATS = average trade size; $|OB|$ = absolute order imbalance.

Table 3: Number of shares traded, realized volatility and its two components.

	RV		$C9999$		$J9999$
	(1)	(3)	(2)	(4)	(5)
$\overline{\phi}$	0.110	0.176	0.106	0.179	-0.042
$\overline{\sigma_\phi}$	0.022	0.033	0.021	0.032	0.013
% *p-value*	0.050	0.220	0.052	0.132	8.872
% + significant	100	99	100	100	2
% − significant	0	0	0	0	72
% Adjusted R^2	40.218	22.484	41.456	23.035	0.522

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} RV_{it-j} + \phi_i V_{it} + \nu_{it} \ (1)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} C9999_{it-j} + \phi_i V_{it} + \nu_{it} \ (2)$$

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \phi_i V_{it} + \nu_{it} \ (3)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \phi_i V_{it} + \nu_{it} \ (4)$$

$$J9999^*_{it} = \alpha_i + \alpha_{im} M_t + \phi_i V_{it} + \nu_{it}, \text{ where } J9999 = max(0, J9999^*) \ (5)$$

The above regressions are run for the 100 largest stocks traded on the NYSE over the period January 1, 1995-September 30, 1999. RV_{it} / $C9999_{it}$ / $J9999_{it}$ is the realized variance / the continuous component of realized variance / the jump component of realized variance of stock i on day t, all multiplied by 10,000. M_t is a Monday dummy, V_{it} is the number of shares traded (divided by 100,000) for stock i on day t and ρ_{ij} measures the persistence of volatility shocks at lag j. Regressions (1) and (2) are estimated by OLS while regressions (3) and (4) are estimated by GMM. Regression (5) is estimated by maximum likelihood. For brevity, we only report the equally-weighted cross-sectional mean coefficient for number of shares traded, with corresponding statistics. Newey-West standard errors and two-sided p-values across stocks are computed. We report the percentage of positive and negative $\hat{\phi}_i$ coefficients which are statistically different from zero at the 5% level. The last row reports the mean adjusted R^2's.

Table 4: Number of trades, realized volatility and its two components.

	RV		$C9999$		$J9999$
	(1)	(3)	(2)	(4)	(5)
$\overline{\beta}$	0.318	0.387	0.315	0.397	-0.137
$\overline{\sigma_\beta}$	0.051	0.055	0.049	0.054	-0.035
% *p-value*	0.020	0.017	0.052	0.090	3.104
% + significant	100	100	100	100	3
% − significant	0	0	0	0	86
% Adjusted R^2	42.334	29.500	43.979	31.654	0.967

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} RV_{it-j} + \beta_i NT_{it} + \nu_{it} \ (1)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} C9999_{it-j} + \beta_i NT_{it} + \nu_{it} \ (2)$$

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \beta_i NT_{it} + \nu_{it} \ (3)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \beta_i NT_{it} + \nu_{it} \ (4)$$

$$J9999^*_{it} = \alpha_i + \alpha_{im} M_t + \beta_i NT_{it} + \nu_{it}, \text{ where } J9999 = max(0, J9999^*) \ (5)$$

The above regressions are run for the 100 largest stocks traded on the NYSE over the period January 1, 1995-September 30, 1999. RV_{it} / $C9999_{it}$ / $J9999_{it}$ is the realized variance / the continuous component of realized variance / the jump component of realized variance of stock i on day t, all multiplied by 10,000. M_t is a Monday dummy, NT_{it} is the number of trades (divided by 100) for stock i on day t and ρ_{ij} measures the persistence of volatility shocks at lag j. Regressions (1) and (2) are estimated by OLS while regressions (3) and (4) are estimated by GMM. Regression (5) is estimated by maximum likelihood. For brevity, we only report the equally-weighted cross-sectional mean coefficient for number of trades, with corresponding statistics. Newey-West standard errors and two-sided p-values across stocks are computed. We also report the percentage of positive and negative $\hat{\beta}_i$ coefficients which are statistically different from zero at the 5% level. The last row indicates the mean adjusted R^2's.

Table 5: Average trade size, realized volatility and its two components.

	RV		$C9999$		$J9999$
	(1)	(3)	(2)	(4)	(5)
$\overline{\gamma}$	0.212	0.100	0.192	0.048	0.125
$\overline{\sigma_\gamma}$	0.124	0.410	0.119	0.404	0.151
% *p-value*	25.075	23.686	27.318	22.400	24.190
% + significant	40	26	37	25	34
% − significant	5	32	7	33	10
% Adjusted R^2	32.443	2.838	33.816	3.219	0.463

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} RV_{it-j} + \gamma_i ATS_{it} + \nu_{it} \quad (1)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} C9999_{it-j} + \gamma_i ATS_{it} + \nu_{it} \quad (2)$$

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \gamma_i ATS_{it} + \nu_{it} \quad (3)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \gamma_i ATS_{it} + \nu_{it} \quad (4)$$

$$J9999^*_{it} = \alpha_i + \alpha_{im} M_t + \gamma_i ATS_{it} + \nu_{it}, \text{ where } J9999 = max(0, J9999^*) \quad (5)$$

The above regressions are run for the 100 largest stocks traded on the NYSE over the period January 1, 1995-September 30, 1999. RV_{it} / $C9999_{it}$ / $J9999_{it}$ is the realized variance / the continuous component of realized variance / the jump component of realized variance of stock i on day t, all multiplied by 10,000. M_t is a Monday dummy, ATS_{it} is the average trade size (divided by 1,000) for stock i on day t and ρ_{ij} measures the persistence of volatility shocks at lag j. Regressions (1) and (2) are estimated by OLS while regressions (3) and (4) are estimated by GMM. Regression (5) is estimated by maximum likelihood. For brevity, we only report the equally-weighted cross-sectional mean coefficient for average trade size, with corresponding statistics. Newey-West standard errors and two-sided p-values across stocks are computed. We also report the percentage of positive and negative $\hat{\gamma}_i$ coefficients which are statistically different from zero at the 5% level. The last row indicates the mean adjusted R^2's.

Table 6: Average trade size, number of trades, realized volatility and its two components.

	RV		*C9999*		*J9999*
	(1)	(3)	(2)	(4)	(5)
$\overline{\gamma}$	0.268	0.407	0.238	0.362	0.096
$\overline{\sigma_\gamma}$	0.131	0.377	0.125	0.369	0.155
% *p-value*	22.331	31.397	25.330	29.575	30.071
% + significant	45	27	38	25	24
% − significant	8	12	8	18	8
$\overline{\beta}$	0.323	0.388	0.319	0.397	-0.134
$\overline{\sigma_\beta}$	0.052	0.059	0.049	0.057	0.032
% *p-value*	0.019	0.047	0.049	0.102	3.266
% + significant	100	100	100	100	3
% − significant	0	0	0	0	85
% Adjusted R^2	42.760	28.704	44.367	30.607	1.032

$$RV_{it} = \alpha_i + \alpha_{im}M_t + \sum_{j=1}^{12} \rho_{ij} RV_{it-j} + \gamma_i ATS_{it} + \beta_i NT_{it} + \nu_{it} \ (1)$$

$$C9999_{it} = \alpha_i + \alpha_{im}M_t + \sum_{j=1}^{12} \rho_{ij} C9999_{it-j} + \gamma_i ATS_{it} + \beta_i NT_{it} + \nu_{it} \ (2)$$

$$RV_{it} = \alpha_i + \alpha_{im}M_t + \gamma_i ATS_{it} + \beta_i NT_{it} + \nu_{it} \ (3)$$

$$C9999_{it} = \alpha_i + \alpha_{im}M_t + \gamma_i ATS_{it} + \beta_i NT_{it} + \nu_{it} \ (4)$$

$$J9999^*_{it} = \alpha_i + \alpha_{im}M_t + \gamma_i ATS_{it} + \beta_i NT_{it} + \nu_{it}, \text{ where } J9999 = max(0, J9999^*) \ (5)$$

The above regressions are run for the 100 largest stocks traded on the NYSE over the period January 1, 1995-September 30, 1999. RV_{it} / $C9999_{it}$ / $J9999_{it}$ is the realized variance / the continuous component of realized variance / the jump component of realized variance of stock i on day t, all multiplied by 10,000. M_t is a Monday dummy, ATS_{it} is the average trade size (divided by 1,000), NT_{it} is the number of trades (divided by 100) for stock i on day t and ρ_{ij} measures the persistence of volatility shocks at lag j. Regressions (1) and (2) are estimated by OLS while regressions (3) and (4) are estimated by GMM. Regression (5) is estimated by maximum likelihood. For brevity, we only report the equally-weighted cross-sectional mean coefficient for average trade size and number of trades, with corresponding statistics. Newey-West standard errors and two-sided p-values across stocks are computed. We also report the percentage of positive and negative $\hat{\gamma}_i$ and $\hat{\beta}_i$ coefficients which are statistically different from zero at the 5% level. The last row indicates the mean adjusted R^2's.

Table 7: Absolute order imbalance, realized volatility and its two components.

	RV		*C*9999		*J*9999
	(1)	(3)	(2)	(4)	(5)
$\overline{\lambda}$	0.435	1.688	0.420	0.690	-0.165
$\overline{\sigma_{\lambda}}$	0.162	0.611	0.160	0.304	0.111
% *p-value*	5.667	9.362	6.854	15.89	16.943
% + significant	83	79	82	63	15
% − significant	0	1	0	6	41
% Adjusted R^2	34.133	2.255	35.524	2.759	0.660

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \textstyle\sum_{j=1}^{12} \rho_{ij} RV_{it-j} + \lambda_i |OB|_{it} + \nu_{it} \ (1)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \textstyle\sum_{j=1}^{12} \rho_{ij} C9999_{it-j} + \lambda_i |OB|_{it} + \nu_{it} \ (2)$$

$$RV_{it} = \alpha_i + \alpha_{im} M_t + \lambda_i |OB|_{it} + \nu_{it} \ (3)$$

$$C9999_{it} = \alpha_i + \alpha_{im} M_t + \lambda_i |OB|_{it} + \nu_{it} \ (4)$$

$$J9999^{*}_{it} = \alpha_i + \alpha_{im} M_t + \lambda_i |OB|_{it} + \nu_{it}, \text{ where } J9999 = max(0, J9999^{*}) \ (5)$$

The above regressions are run for the 100 largest stocks traded on the NYSE over the period January 1, 1995-September 30, 1999. RV_{it} / $C9999_{it}$ / $J9999_{it}$ is the realized variance / the continuous component of realized variance / the jump component of realized variance of stock i on day t, all multiplied by 10,000. M_t is a Monday dummy, $|OB|_{it}$ is the absolute order imbalance (divided by 100) for stock i on day t and ρ_{ij} measures the persistence of volatility shocks at lag j. Regressions (1) and (2) are estimated by OLS while regressions (3) and (4) are estimated by GMM. Regression (5) is estimated by maximum likelihood. For brevity, we only report the equally-weighted cross-sectional mean coefficient for absolute order imbalance, with corresponding statistics. Newey-West standard errors and two-sided p-values across stocks are computed. We also report the percentage of positive and negative $\hat{\lambda}_i$ coefficients which are statistically different from zero at the 5% level. The last row indicates the mean adjusted R^2's.

Table 8: Absolute order imbalance, number of trades, realized volatility and its two components.

	RV		$C9999$		$J9999$
	(1)	(3)	(2)	(4)	(5)
$\overline{\lambda}$	-0.163	-0.753	-0.177	-0.762	0.188
$\overline{\sigma_\lambda}$	0.164	1.346	0.160	1.188	0.136
% *p-value*	28.761	31.275	27.853	28.967	26.879
% + significant	1	1	1	1	41
% − significant	42	40	45	41	0
$\overline{\beta}$	0.347	0.483	0.347	0.500	-0.164
$\overline{\sigma_\beta}$	0.052	0.150	0.049	0.135	0.036
% *p-value*	0.049	3.339	0.056	2.423	4.130
% + significant	100	94	100	95	0
% − significant	0	0	0	0	88
% Adjusted R^2	42.898	17.318	44.603	20.897	1.233

$$RV_{it} = \alpha_i + \alpha_{im}M_t + \sum_{j=1}^{12} \rho_{ij} RV_{it-j} + \lambda_i|OB|_{it} + \beta_i NT_{it} + \nu_{it} \ (1)$$

$$C9999_{it} = \alpha_i + \alpha_{im}M_t + \sum_{j=1}^{12} \rho_{ij} C9999_{it-j} + \lambda_i|OB|_{it} + \beta_i NT_{it} + \nu_{it} \ (2)$$

$$RV_{it} = \alpha_i + \alpha_{im}M_t + \lambda_i|OB|_{it} + \beta_i NT_{it} + \nu_{it} \ (3)$$

$$C9999_{it} = \alpha_i + \alpha_{im}M_t + \lambda_i|OB|_{it} + \beta_i NT_{it} + \nu_{it} \ (4)$$

$$J9999^*_{it} = \alpha_i + \alpha_{im}M_t + \lambda_i|OB|_{it} + \beta_i NT_{it} + \nu_{it}, \text{ where } J9999 = max(0, J9999^*) \ (5)$$

The above regressions are run for the 100 largest stocks traded on the NYSE over the period January 1, 1995-September 30, 1999. RV_{it} / $C9999_{it}$ / $J9999_{it}$ is the realized variance / the continuous component of realized variance / the jump component of realized variance of stock i on day t, all multiplied by 10,000. M_t is a Monday dummy, $|OB|_{it}$ is the absolute order imbalance (divided by 100), NT_{it} is the number of trades (divided by 100) for stock i on day t and ρ_{ij} measures the persistence of volatility shocks at lag j. Regressions (1) and (2) are estimated by OLS while regressions (3) and (4) are estimated by GMM. Regression (5) is estimated by maximum likelihood. For brevity, we only report the equally-weighted cross-sectional mean coefficient for average trade size and number of trades, with corresponding statistics. Newey-West standard errors and two-sided p-values across stocks are computed. We also report the percentage of positive and negative $\hat{\lambda}_i$ and $\hat{\beta}_i$ coefficients which are statistically different from zero at the 5% level. The last row indicates the mean adjusted R^2's.