THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

February 18, 2008

Wp# 0039MSS-045-2008

Is Trading Imbalance a Better Explanatory Factor in the Volatility Process? Intraday and Daily Evidence from E-mini S&P 500 Index Futures and Information-Based Hypotheses

An-Sing Chen
(National Chung Cheng University, Taiwan)

Hui-Jyuan Gao
(National Chung Cheng University, Taiwan)

Mark Leung
(University of Texas at San Antonio)

Department of Management Science & Statistics,
University of Texas at San Antonio,
San Antonio, TX 78249, U.S.A

UTSA COLLEGE OF BUSINESS KNOWLEDGE FOR A NEW WORLD

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

Is Trading Imbalance a Better Explanatory Factor in the Volatility Process? Intraday and Daily Evidence from E-mini S&P 500 Index Futures and Information-Based Hypotheses

Abstract

This paper examines trading imbalance as well as traditional trading variables in the volume-volatility relation in futures market. Unlike the majority of studies which utilize daily data, our empirical investigation compares an array of intraday frequencies (from five minutes to one hour) with daily interval. The primary analysis is conducted through a series of GARCH tests and the findings are then confirmed by a set of two-stage least square regressions. Since this paper adopts an information-based framework to explain the volume-volatility relation, unexpected trading variables are used to proxy for new market information. Results indicate that different trading imbalance metrics are useful and more significant than traditional trading variables in explaining the volatility relation for all daily and intraday intervals. Empirical findings support the existence of asymmetric information hypothesis at all intervals. On the other hand, mixture of distributions and difference in opinion hypotheses are validated in only some intraday intervals. Moreover, not only are the conclusions from daily observations not the same as the ones from intraday counterparts but also there are differences in the results between longer and shorter intraday intervals.

Keywords: Futures markets, price volatility, trading imbalance, number and volume of trades, asymmetric information, difference in opinion, mixture of distributions, GARCH and persistence effect.

JEL Code: G1, G13

Acknowledgment

This research was partially funded by the COB grant made available by the University of Texas. The authors wish to thank the valuable comments from the anonymous referees and the constructive suggestions from Hazem Daouk of Cornell University. Any errors in this paper are of our own.

Is Trading Imbalance a Better Explanatory Factor in the Volatility Process? Intraday and Daily Evidence from E-mini S&P 500 Index Futures and Information-Based Hypotheses

1. Introduction

Price volatility is a fundamental element in financial research. Contingent claims pricing, risk management, asset allocation, market efficiency, and many other avenues of studies use volatility as a basic building block. Therefore, volatility has been widely recognized as one of the factors contributing profound implications in finance. A large number of these studies focus on the association between price volatility and trading volume. According to the information proposition, new information release causes transactions, which in turn conveys information to market participants and generates price change. In other words, information release usually creates heavier trading volume and more volatility shocks.

Direct quantification of trading activities can be a noisy measure of information because trading occurs not only at the time information arrives but also when investors possess diverse opinions or interpretations of news release. Hence, trading volume, which can be decomposed into the number and size of trades, as a proxy measure for information is not always appropriate.

There is an ample collection of empirical research studies investigating how trading volume and volatility are related. Unfortunately, most studies focus on the number and size of trades and ignore a crucial element of market microstructure models (as suggested by Kyle (1985) and others), that is, price volatility being affected by order imbalance. Market makers often infer information from order imbalance and then upwardly revise the price when there are excessive buy orders. The vice versa scenario also holds. Such behavior is supported by many empirical studies (like Glosten and Harris (1988), Madhavan et al. (1997), Huang and Stoll (1997), Chodia, Roll and Subrahmanyam (2002)). Given order imbalance is useful in explaining movements in price and changes in quote, it should also play an important role in the process of price volatility. In addition, if most insiders are confident of the information, their orders will cluster on one side of the market and subsequently induce a change in price. Thus, the extent of trading imbalance should reflect the quality of private information and affect return volatility.

A review of existing literature suggests that, although there is a collection of models explaining this volume-volatility relation, they primarily fall into one of the three major theoretical frameworks - mixture of distributions hypothesis, difference in opinion hypothesis, and asymmetric information hypothesis. For the sake of brevity, interested readers should refer to the Appendix for a detailed explanation of these information-based hypotheses and a comprehensive review of the literature.

Although the work in addressing the volume-volatility relation is abundant, there exist very few papers discussing the connection between order imbalance and volatility in trading. Chan and Fong (2000) is one of the few studies examining the roles of number of trades, size of trades and order imbalance in volume-volatility relation. Based on daily observations from NYSE and NASDAQ stock markets, they find that size of trade provides better information than number of trades in the relation. They also discover that order imbalance explains a substantial portion of daily price movements.

Therefore, in order to examine the three information-based hypotheses for volume-volatility relation, this paper adopts number of trades, trading volume, and trading imbalance as experimental factors. For theoretical and practical reasons, two trading imbalance metrics are proposed. One represents the number of trading imbalance and the other is the volume of trading imbalance. We contend that the information content of trades may not be fully captured by number of trades and trading volume alone. Instead, trading imbalance should contain information about the degree of information asymmetry, which is not revealed directly from the traditional trading variables. One of our motivations is thus to explore whether trading imbalance plays an important role in explaining the volatility process. Besides, it is interesting to find out if these trading variables exert different levels of influence on persistence effect and on the validity of the three prevailing hypotheses for volatility.

Particularly in this study, we extend the findings from previous studies and create a more coherent group of tests encompassing both existing and new issues. First**,** with the notion that size of trade is likely to be positively related to the quality of information, Chan and Fong (2000) state that the asymmetric information hypothesis is supported. At the same time, they also find the number of trades affects volatility, implying that the mixture of distributions hypothesis holds. Nonetheless, their study does not compare the relative degrees of influence by number of trade, size of trade, and order imbalance on volatility. In this paper, we abridge this gap by looking at possible dominance by any of the three information-based hypotheses over a range of conditions.

Second, from an information standpoint, it is more appropriate to classify trading variables into expected and unexpected categories. A trading variable generated from normal market activity, conditional on past values, is called an expected trading variable. When a trading variable is derived by information unpredictable by the market, it is called an unexpected trading variable. If a trading variable is used as the proxy for new information flow into the market, we can probably observe the relation between the unexpected trading variable and volatility. A number of studies have supported this concept [1]. Since we explain volume-volatility relation from an information-based perspective, unexpected trading variables are the primary focus in our experiment[2].

Third, E-mini futures contracts can be traded around the clock on the electronic GLOBEX trading system. Along with their smaller sizes, this family of financial instruments has been expanding rapidly since its introduction by CME in 1997. Hasbrouck (2003) demonstrates that the largest informational contributions arise from the electronically traded futures contracts, and finds most of price discovery in the E-mini futures market. Consistent with Hasbrouck (2003), Kurov and Lasser (2004) examine price dynamics in the regular and the E-mini futures markets. They suggest that the E-mini market is an important satellite market. Given these notions, we believe that it is insightful to explore the volume-volatility relation via E-mini S&P 500, the first product introduced to the E-mini futures market.

Fourth, a persistent effect of volatility has been found in GARCH models. Lamoureax and Lastrapes (1990) state that trading volume can be a good proxy for arrival of information to the market and for explaining the persistence of return volatility of individual shares. Nevertheless, in their empirical results, the introduction of trading volume does not eliminate the GARCH persistence. In order to distinguish which trading variables play a crucial role in volatility model, we test empirically the reduction in the degree of GARCH volatility persistence. Simply put, if trading imbalance plays a significant role in volatility, we will find the coefficient of trading imbalance significant. Also, the persistence of volatility should be substantially reduced when trading imbalance is added to the volume-volatility model. In the auxiliary experiment, we employ two-stage least square regressions to confirm these conjectures based on GARCH models[3].

Fifth, information shock to the market should not persist for a long duration if there is a considerable degree of market efficiency. Taking this into consideration, we conduct testing with respect to both daily and intraday data at different frequencies, including hourly, 30-minute, 15-minute, and 5-minute intervals. Unlike former studies which are solely based on daily observations, our study provides an

---

1 See Section 2.2 for more detailed information.

2 Expected trading variables as well as variables based on total values are also empirically tested. Results are generally similar to the ones obtained by unexpected trading variables.

3 Separate GRACH tests are conducted using corresponding trading variables in terms of total values and unexpected components (details will be explained in the next section.)

opportunity to understand whether or not the length of interval frequency is a non-trivial factor in volatility estimation. In other words, we can find out if various trading variables behave differently at different frequency intervals, especially whether findings from daily observations are applicable to intraday observations.

Our paper is organized as follows. In the next section, we describe the experimental models and methodologies. Details of data preparation including the construction of our proposed trading imbalance metrics and computation of the unexpected components of trading variables are also provided. Theoretical foundation as well as the connections among the three information-based hypotheses and the two sets of empirical models (GARCH and two-stage least square) are explained. Section 3 describes the sample data and reports descriptive statistics for the trading variables. Empirical results of our primary experiment are presented and discussed in Section 4. Auxiliary robustness check of the experimental conclusions is conducted in Section 5. Section 6 summarizes and concludes the paper.

2. Experimental Metrics and Models

2.1 Data preparation and trading imbalance metrics

E-mini S&P 500 index futures are examined in our empirical investigation. The data are obtained from ANFutures, which outsources the data directly from CME. The ANFutures database contains intraday information on contract symbol, trade date, trade time to the nearest second, trade volume, adjusted price of trade[4], tick number and tick volume, and other related records.

Besides using the observed number of trade and trading volume as proxies for information content, trading imbalance metrics are computed from tick data Up tick is defined as a tick whose value is higher than or the same as the previous one. Conversely, a down tick is defined as one with value lower than the previous one. Our detailed intraday dataset allows us to construct two types of trading imbalance metrics, one based on the *numbers* of up and down ticks in a prescribed time interval and another based on the *volumes* of up and down ticks[5].

The metric based on volume measures information content of the volume of contracts traded as well as of the frequency of trades within a particular time interval.

---

4 If you have an open position and its expiration date is near, it is possible to roll over to the next active month in order to avoid delivery obligations. In the ANFutures database, rollovers for the E-mini S&P 500 index futures contracts take place 10 days prior to expiration and the individual futures contracts are spliced together from “Roll-Over-Day” to “Roll-Over-Day” to form one continuous time series (continued futures contract).

5 Number of trade is computed as the sum of up-tick-number and down-tick-number. Trading volume is computed as the sum of up-tick-volume and down-tick-volume. The first imbalance metric is computed from the absolute value of the up-tick-number minus the down-tick-number; similar, the second imbalance metric is computed from the absolute value of the up-tick-volume minus the down-tick-volume. All four trading variables are presented in thousands.

If there exists any extra information in the volume of contracts traded, these metrics should pick it up. The metric based on number, on the other hand, does not contain this information and hence may create the subtle effect of assigning relatively heavier weights on trades with smaller contract volumes.

It is observed that the four trading variables increase gradually throughout the sample period. Therefore, it is necessary to remove the deterministic trend to avoid spurious results. To de-trend trading variables, we adopt the following family of regression models with linear and nonlinear time trends, where V_t represents number of trade, trading volume, or one of the two trading imbalances, respectively:

$$V_t = c + a_1 T + a_2 T^2 + \varepsilon_t \qquad (1)$$

The residual terms from regressions are treated as the de-trended number of trade, the de-trended trading volume and the de-trended trading imbalances. All regressions show that all estimated coefficients a_1 and a_2 are significantly different from zero. Readers should be aware that all trading variables in this paper are de-trended.

In the empirical tests, we examine volume-volatility and imbalance-volatility relations to reveal and thus confirm the underlying hypotheses and their implications. As explained earlier, the three hypotheses are the mixture of distributions hypothesis, the difference in opinion hypothesis, and the asymmetric information hypothesis. Moreover, we compare the roles of volume and imbalance in volatility process and find out which variable, if any, captures information better.

In order to assess these three information-based hypotheses, GARCH model and two-stage least square regression are employed in our experiment. Technically, we observe the magnitude of decline in GARCH model's persistence effect to understand the roles of volume and trading imbalances on conditional variances. On the other hand, we use the residual of two-stage least square regression as proxy for volatility, allowing us to re-examine the relations of volume-volatility and imbalance-volatility. Further, we may uncover if there exists a dominant hypothesis across different frequency intervals.

2.2 Unexpected components of number of trade, trading volume, and trading imbalances

A handful of studies in the literature focuses on trading volume. Representative examples are Bessembinder and Seguin (1992 and 1993), Daigler and Wiley (1999), Lee and Rui (2002), and Arago and Nieto (2005). These studies find that unexpected volume shocks have a larger effect on volatility. Therefore, in this paper, the unexpected components of number of trade, trading volume, and trading imbalances (all de-trended) are used to proxy for new information content in the market.

The model used to estimate the various unexpected components is as follows:

$$Y_t = c + \sum_{i=1}^{10} b_1 \left| r_{t-j} \right| + \sum_{j=1}^{10} Y_{t-j} + \sum DW_t + \varepsilon_t \qquad (2)$$

Y_t is number of trade, trading volume, or trading imbalances (all de-trended). $\left| r_{t-j} \right|$ is the absolute value of price change.[6] DW is the dummy variable controlling the day of week for daily data or the time of open and close for intraday data. The residual ε_t from regression is the unexpected component of the respective trading variable. Henceforth, the experimental variables Num, $Volm$, I^{num}, and I^{volm} denote the unexpected components of number of trade, trading volume, and the two trading imbalances, respectively.

2.3 Examination of volatility persistence based on GARCH models

Lamoureux and Lastrapes (1990) document substantial reduction in volatility persistence when trading volume is included in the variance equation of the GARCH(1,1) model. Stemming from their finding, we should find a larger reduction of persistent effect if trading imbalance captures more information than volume does, i.e., if trading volume in the conditional variance equation is replaced by trading imbalance.

Mathematical specification of the GARCH(1,1) model in our test is shown below. The conditional mean regressions are expressed in Equations (3-1) for daily interval and (3-2) for various intraday intervals, respectively.

Mean equations:

$$R_t = \sum_{k=1}^{5} \phi_{1k} D_k + \sum_{j=1}^{12} \phi_{2j} R_{t-j} + \varepsilon_t \qquad \varepsilon_t \left| \Omega_{t-1} \sim N(0, \sigma_t^2) \right. \quad \text{(3-1)} \quad \text{for daily}$$

$$R_t = \phi_0 + \phi_{11} OPEN_t + \phi_{12} CLOSE_t + \sum_{j=1}^{12} \phi_{2j} R_{t-j} + \varepsilon_t \qquad \text{(3-2)} \qquad \text{for intraday}$$

Variance equations:

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 \qquad (4)$$

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_1 \cdot Num_t \qquad \text{(4-1)}$$

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_2 \cdot Volm_t \qquad \text{(4-2)}$$

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_3 \cdot I_t^{num} \qquad \text{(4-3)}$$

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_4 \cdot I_t^{volm} \qquad \text{(4-4)}$$

[6] Numerous studies (such as Schwert (1990) and Gallant, Rossi and Tauchen (1990), Bessembinder and Sequin (1993)) provide evidence that past volatilities have predictive power for forecasting volumes. To capture this power, absolute price change is used as proxy for volatility in regression. The volatility proxy is also used by Watanabe (2001).

where R_t is the change of natural logarithm of futures price at interval t. D_k represents the five dummy variables for the days of week. R_{t-j} is the lagged term used to control for any serial dependence in return. Dummy variables OPEN and CLOSE are included in the intraday conditional mean equations[7].

In variance equations, coefficients α and β reflect the dependence of the current volatility upon its past levels, including information about volatility during the previous period (ε_{t-1}^2) and fitted variance from the model during the previous period (σ_{t-1}^2). The sum of ($\alpha + \beta$) indicates the degree of volatility persistence. Number of trades, trading volume and the two trading imbalances are added to the base-case Equation (4) as explanatory factors. They are shown in conditional variance Equations (4-1) to (4-4).

Moreover, we evaluate the three information-based hypotheses with respect to the sign of estimated coefficient γ_i. The expected signs of the coefficients in variance equations are displayed in Table 1:

Table 1 Proper signs of coefficients and comparison of hypotheses in variance equation

Basis of metric	Supported Hypothesis	γ_1	γ_2	γ_3	γ_4
Tick number	**Mixture of distributions hypothesis**	+			
	Asymmetric information hypothesis			+	
	Both M.D.H and A.I.H	+		+	
Tick volume	**Difference in opinion hypothesis**	+			
	Asymmetric information hypothesis				+
	Both of D.O.H and A.I.H	+			+

There are three empirical rationales from these variance equations. First, the significance of the coefficient γ_i provides evidence on whether different information influences volatility in the presence of conditional heteroscedasiticity in return. A significantly positive coefficient indicates the usefulness of the respective trading variable and, thereby, the validity of the corresponding hypothesis. Second, if the trading variables in conditional variance regression are well fitted, the GARCH persistent effect, defined earlier as ($\alpha + \beta$), should be reduced. Third, if trading imbalances (I^{num}, I^{volm}) are better factors than number of trade (Num) or trading volume ($Volm$) in explaining volatility, the degree of reduction in volatility persistence should be larger in Equations (4-3 and 4-4) than in Equations (4-1 and

[7] In order to include the first and the last intervals for all frequencies, we have extended observation time period from 8:30am to 15:30 pm for the 30-minute frequency, and from 8:00 am to 16:00 pm for the hourly frequency.

4-2).

Further, to understand if there is a dominant hypothesis for explaining volume-volatility relation (e.g., whether trading imbalance is better than number of trade or trading volume as indicator for information content), we include number of trade and trading imbalance simultaneously in the conditional variance regressions (5-1 and 5-2). Using Equation (5-1) based on tick number, we examine variables Num and I^{num} to compare the fitness of the mixture of distribution hypothesis with that of the asymmetric information hypothesis. Similarly, using Equation (5-2) based on tick volume, we compare the fitness of the difference in opinion hypothesis ($Volm$) with that of the asymmetric information hypothesis. (I^{volm}).

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_1 \cdot Num_t + \gamma_3 \cdot I_t^{num} \quad (5\text{-}1)$$

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_2 \cdot Volm_t + \gamma_4 \cdot I_t^{volm} \quad (5\text{-}2)$$

2.4 Evaluation of volatility based on residuals of two-stage least square regressions

From the work by Schwert (1990), Jones, Kaul, and Lipson (1994), Chan and Fong (2000), and Chordia, Sarkar, and Subrahmanyam (2005), volatility can be estimated by the absolute residuals of regressions (see Equation (3-1) for daily interval and Equation (3-2) for intraday intervals.)

To examine the daily volume-volatility relation, the specification of volatility regression is as follows:

$$|\varepsilon_t| = \gamma_0 + \sum_{i=1}^{4} \phi_{1k} D_k + \sum_{j=1}^{12} \delta_j |\varepsilon_{t-j}| + \gamma_1 Trade_t + \eta_t \qquad (6\text{-}1)$$

Daily dummies D_{kt} (k=1 to 5) in the volatility regression are used to capture differences in volatility by the day of week. *Trade* is a proxy for one of the four unexpected trading variables (Num_t, $Volm_t$, I_t^{num}, and I_t^{volm}), depending on the subject of analysis.

For intraday volatility regression, we employ the following general model:

$$|\varepsilon_t| = \gamma_0 + \phi_1 OPEN_t + \phi_2 CLOSE_t + \sum_{j=1}^{12} \delta_j |\varepsilon_{t-j}| + \gamma_1 Trade_t + \eta_t \qquad (6\text{-}2)$$

Wood et al. (1985) and Jain and Joh (1988) document that trading volume follows a U-shaped intraday pattern. Foster and Viswanathan (1993) show that volatility at market opening is much higher because of the asymmetric information effect arising from no overnight trading. Therefore, we insert dummy variables OPEN and CLOSE in the intraday volatility regressions to capture the possible asymmetric effect.

Similar to the GARCH model described in the last section, we use the sign of coefficient in Equation (6-1) to determine which hypothesis(es) is(are) valid in the market. The expected results for various hypotheses have been summarized in Table 1. In addition to the GARCH persistent effect, model explanatory power (R^2) are used to verify whether trading imbalance is superior to volume in capturing more information. Essentially, the higher the explanatory power (R^2), the better the capacity of that variable in capturing volatility information.

3. Data set description and preliminary analysis

3.1 Sample

The E-mini S&P 500 futures were introduced by the Chicago Mercantile Exchange (CME) in September 1997. Compared to the regular S&P 500 index futures, E-mini futures are 1/5 of the size of the regular counterparts. They not only enable private investors to participate and trade the instruments via internet but also provide the industry with its first small-order electronic order routing and execution system. The E-mini S&P 500 futures are exclusively traded on GLOBEX (Chicago Mercantile Exchange's platform) almost 24 hours a day.

A typical 24-hour trading day is separated into regular trading hours (when the spot market is also open) and non-regular trading hours (when only E-mini futures are traded). Since the tick number (or volume) for E-mini futures before the opening and after the closure of the floor-traded futures markets is relatively small, our empirical analysis focuses on the regular trading time for E-mini futures (i.e., the regular trading hours of floor-traded futures) only. The regular trading time of the S&P 500 index runs from 8:30 to 15:15.

The data for our analysis cover the period from April 2, 1998 to March 9, 2005 and are compiled for different frequency intervals - 5-minute, 15-minute, 30-minute, hourly, and daily. In the ANFutures database, rollovers for the E-mini S&P 500 index futures take place 10 days prior to expiration and the individual futures contracts are spliced together from "Roll-Over-Day" to "Roll-Over-Day" to form one continuous time series (continued futures contract).

3.2 Descriptive statistics and correlations

Table 2 presents the descriptive statistics for return, volatility, number of trade, trading volume, number of trading imbalance, and volume of trading imbalance with respect to different frequency intervals. Returns generally show significantly negative skewness with the only exception in the case of hourly frequency. The absolute residuals (proxy for volatility measured by two-stage least square regression) indicate a high level of positive skewness and are leptokurtic. All trading variables presented in the table have been de-trended using the procedure explained before. Compared with traditional trading variables (Num and $Volm$), trading imbalance metrics (I^{num} and I^{volm}) have lower standard deviations in both tick number and volume. The magnitude increases when frequency gets shorter, except for daily interval. For example, in the five-minute interval, standard deviation is 0.19 for Num, and only 0.06 for I^{num}; 3.10 for $Volm$ and only 1.12 for I^{volm}. In the hourly interval, standard deviation is 2.85 for Num, and only 0.20 for I^{num}; 25.58 for $Volm$ and only 4.31 for I^{volm}. It seems that number of trade and trading volume are noisier measures than trading imbalances, especially in intraday intervals. The absolute residuals for all trading variables are right-skewed and highly leptokurtic in all frequencies.

Table 3 documents the correlations among volatility, the two traditional trading variables (number of trade and trading volume), and the two trading imbalance metrics. Results are tabulated in Panels A to E for various frequencies. For tick number metric, the correlation between volatility and number of trading imbalance (I^{num}) is higher than that between volatility and number of trades (Num). For tick volume metric, the correlation between volatility and volume of trading imbalance (I^{volm}) is also higher than that between volatility and trading volume ($Volm$). At this preliminary stage, it seems that trading imbalance metrics play stronger roles than number of trade and trading volume in capturing and explaining the volatility fluctuation process.

4. GARCH tests for persistent effect and information-based hypotheses

4.1 Persistent effect in number of trade, trading volume and trading imbalance

From an information point of view, it is more proper to observe the unexpected components (explained in Section 2.2) of trading variables. We gauge the degree of reduction in volatility persistence in GARCH model to determine whether unexpected number of trades (Num), unexpected trading volume ($Volm$) and the two unexpected trading imbalance metrics (I^{num} and I^{volm}) play significantly different roles in volatility[8]. According to Lamoureux and Lastrapes (1990), the time-varying pattern of conditional volatility may be generated by serial correlation in the information arrival process. As a result, the conditional variance displays patterns of time dependence (or clustering). Empirically, this implies that, when a proxy for information flow is inserted into the conditional variance equation,

---

[8] The total values and expected values of number of trades, trading volume and trading imbalance metrics are also examined in GARCH test. In order to save space, the results are not reported here, but are available from authors on request.

observed volatility persistence will diminish. Simply put, larger reduction in persistence effect should be realized by better proxy variable for information.

Relevant results are shown in Table 4. For intraday intervals, the introduction of *Num* and *Volm* into conditional variance regression reduces the variance persistence ($\alpha+\beta$) marginally (in magnitudes of only 0.01 to 0.03), except for the hourly interval. However, when *Num* and *Volm* are replaced by I^{num} and I^{volm}, the degree of persistence effect decreases significantly. The persistence level of volatility ($\alpha+\beta$) decreases from 0.99 to 0.73 for the hourly interval, from 0.99 to 0.73 for the 30-minute interval, from 0.97 to 0.78 for the 15-minute interval, and from 1.02 to 0.76 for the 5-minute interval. Also, the reduction is larger for tick number trading imbalance than in its tick volume counterpart, except for the 15-minute interval. The observations are consistent with our preliminary finding in Table 3 in that the correlation between volatility and number of trading imbalance (I^{num}) is the highest among trading variables. From Table 4, it can be seen that the unexpected trading imbalance variables significantly reduce the persistence ($\alpha+\beta$) from the base case across all intraday intervals except for the case of volume of trading imbalance (I^{vol}) at the 5-minute interval in which persistence remains quite high[9]. These results echo those of Speight, Mcmillan and Gwilym (2000), who use unexpected volume proxy for information flow.

For daily interval, it can be seen that adding number of trades, volume, or trading imbalances to the conditional variance regression does not substantially reduce the variance persistence ($\alpha+\beta$). The results are similar to the finding from Girma and Mougoue (2002). Their study examines the relation between futures spread volatility, volume, and open interest in daily data. They find that the persistence of volatility is high and that introducing volume only marginally reduces the GARCH effect in volatility. Moreover, Luu and Martens (2003) and Arago and Nieto (2005) apply daily volume and other trading variables to GARCH model to examine volume-volatility relation. Their persistent level is around at 0.98.

In summary, our empirical outcomes suggest that both unexpected trading imbalances perform better than traditional unexpected trading variables in capturing volatility. Also, the substantial reduction in persistence level indicates that trading imbalances are good proxies for information content, especially for the intraday intervals. GARCH tests based on total values and expected components of trading

---

[9] At the 5-minute interval, number of trading imbalance (I^{num}) may be a more sensitive measurement/variable than volume of trading imbalance (I^{vol}).

variables are also performed. Results (not reported) are very similar to the conclusions from the unexpected trading variables.[10]

4.2 GARCH test of contemporaneous volume-volatility relation

The arrival of new information induces a sequence of trades that reveal the pricing implication of unannounced information. The dynamic process of incorporating information into market price simultaneously affects price movement and trading volume. Thus, it is possible to observe a contemporaneous relation between volatility and trading variables.

In addition to the testing of persistent effect ($\alpha + \beta$), our GARCH model can also be used to evaluate the hypothesis(es) supporting the volume-volatility relation. This is done by examining the signs and the significance levels of estimated coefficients in Table 4. Our results indicate that all coefficients of trading variables at different intervals are significantly positive at the 1 % level, an experimental finding consistent with the predictions from information-based hypotheses.

The significantly positive coefficients of number of trades (γ_1) support the mixture of distribution hypothesis. When new information arrives, the observed number of trades, which serves as a signal of information release, will induce people to trade. Thus, a positive relation between number of trades and volatility is found. The outcome of this study is supported by the literature. For example, Jones, Kaul and Lipson (1994) find that the number of trades explains almost all of the variation of volatility at daily level. Wu and Xu (2000) also conclude that the number of trades has a significantly positive relation with volatility at half-hourly intervals.

Next, the significantly positive coefficients of trading volume (γ_2) support the difference in opinion hypothesis across all intervals. This implies that participants in the E-mini futures market are heterogeneous traders in that they show different opinions on information release and trading. The results are in agreement with Kalev, Liu, Phan and Jarnecic (2004), which introduce trading volume to GARCH model and obtain significantly positive coefficient in conditional variance. Bessembinder and Sequin (1993) and Watanabe (2001) also find a significant contemporaneous positive relation between volatility and unexpected volume[11].

---

[10] Reduction in persistence level measured by total value or expected component is less than the reduction in the case of unexpected component. Results are available on request.

[11] When we use total values of trading variables in the GARCH model, the coefficients of number of trade (γ_1) and trading volume (γ_2) in the conditional variance equation are not significant at daily interval. At intraday intervals, the coefficients of number of trade (γ_1) are only significantly positive in the 15-minute and the 5-minute intervals. Form the results, it reconfirms unexpected components (relative to trading variables based on total value) are more appropriate proxies for information.

Since trading imbalance may convey information about the degree of information asymmetry not directly revealed by number of trades or trading volume, we introduce two trading imbalance metrics to assess the information asymmetry hypothesis. From Table 4, the coefficients of trading imbalance metrics (γ_3 and γ_4) are significant. Therefore, the asymmetric information hypothesis is supported. This suggests that informed traders submit orders based on private information, which is reflected by the degree of trading imbalance. The results align with those of Wu and Xu (2000), and Chan and Fong (2000), which claim imbalance variables playing non-trivial roles in volume-volatility relation.

In summary, number of trades, trading volume and trading imbalance metrics are good proxies for information. Each of these variables plays an important role in volatility-volume relation at both daily and intraday levels. As a conclusion, the mixture of distributions, the difference in opinion, and the asymmetric information hypotheses are empirically supported by the volatility-volume relation posed in the E-mini futures market.

4.3 Explanatory power of the three information based hypotheses

Given our experimental results, two questions are immediately raised. Among the examined trading variables, is there any variable playing a relatively more powerful role in volume-volatility relation? Which information-based hypothesis, if any, is more fitted in explaining volume-volatility relation in the futures market? The study by Chan and Fong (2000) lays the ground work on addressing these issues. In this paper, we extend their study in two ways. First, we compare various information-based hypotheses according to number of trades, trading volume and trading imbalance metrics. Our empirical investigation should reveal any variable or hypothesis dominant in explaining volume-volatility relation. Second, while Chan and Fong (2000) examine volatility-volume relation at daily frequency, this study evaluates the three information-based hypotheses under different daily and intraday intervals. The use of multiple frequency intervals should shed some light on the impact of interval length on volatility and futures market efficiency.

In previous sections, it is found that the three information-based hypotheses are valid but, in terms of persistent effect, trading imbalances provide more information than traditional trading variables (i.e., number of trades and trading volume) in explaining the volatility process. In this section, we compare the asymmetric information hypothesis (as represented by trading imbalance) with the mixture of distributions hypothesis (as measured by tick number) and, consequently, with the difference in opinion hypothesis (as measured by tick volume). To facilitate the comparative evaluation, we introduce unexpected number of trades (*Num*) and unexpected number of trading imbalance (I^{num}) simultaneously to the conditional variance regression. If number of trading imbalance is a better explanatory variable than number of trades, significantly positive coefficient will appear only for trading imbalance. Then the asymmetric information hypothesis, relative to the mixture of distributions hypothesis, is a more suitable explanation for volume-volatility relation. Similar logic can be applied to the comparison between the asymmetric information and the difference in opinion hypotheses through an examination of unexpected trading volume (*Volm*) and unexpected volume of trading imbalance (I^{volm}). Results of these regression tests are presented in Tables 5 and 6, respectively.

From Table 5, the coefficients (γ_3) for unexpected number of trading imbalance are significantly positive for all frequency intervals whereas the coefficients (γ_1) for number of trades are significantly positive for the five and the 15-minute intervals only. The difference points out a clear domination by the asymmetric information hypothesis at daily and higher intraday frequency intervals. Nevertheless, both the asymmetric information and the mixture of distributions hypotheses are useful in explaining volume-volatility relation at very short intraday intervals. This finding is possibly linked to the persistence effect associated with the mixture of distributions hypothesis and displayed in very short intraday intervals (see Table 4.)

Table 6 shows that the asymmetric information hypothesis is valid at all daily and intraday intervals as the corresponding coefficients (γ_4) for volume of trading imbalance are positively significant. On the other hand, significantly positive coefficients (γ_2) related to the difference in opinion hypothesis occur only at the five, 15 and 30 minute intervals. Thus, the hypothesis is not able to explain the volatility process on a daily and hourly basis. Together with the results from Table 5, although we find empirical evidence for all three information-based hypotheses at some intraday levels, the asymmetric information hypothesis seems to be more universal. Daily observations support only the asymmetric information hypothesis[12] [13].

There are several implications from these empirical findings. First, traditional trading variables (number of trades and trading volume) are quite "noisy" in

---

[12] The comparative evaluation also had been done based on total values of trading variables, their conclusion are very similar to the ones based on unexpected values. Results are available on request.

[13] The finding is consistent with the Luu and Martens (2003) which examine the S&P 500 index futures market at daily interval. Although they use trading volume (this is a different variable from trading imbalance), their experiment finds that the coefficient of trading volume is significant in the conditional variance model. However, when they include intraday volatility, the coefficient of trading volume is no longer significant.

measuring volatility. The problem becomes more obvious when frequency interval gets longer. Second, the E-mini index futures market is rather efficient because innovations of public information are often adapted in less than one day. Third, the degree of trading imbalance should reflect the quality of private information. Informed traders often derive their informational advantage from order flows and private information that cannot be easily obtained by general investors. This possibly explains why only asymmetric information hypothesis holds for the daily interval.

5. Auxiliary testing based on two-stage least-square regression

To check the robustness of our findings and to strengthen our conclusions, we devise a series of auxiliary tests based on the two-stage regression model outlined by Chan and Fong (2000). In other words, we re-examine the volume-volatility relation using a different approach.

Table 7 reports the estimation results when trading variables are applied individually and in pairs to the regression models. This paradigm follows the analytical framework in the previous GARCH experiments and allows a parallel comparison of the conclusions. For the individual cases, all trading variables are significantly positive at the 1% level at various intervals. These outcomes are generally consistent with the ones shown in Table 4 (see Section 4.1). Moreover, a comparison of R^2 illustrates that the two trading imbalance metrics have higher explanatory power than those of number of trades and trading volume. The increments in R^2 are approximately 4~5% at daily interval and 2~3% at intraday intervals. Therefore, the results are in alignment with previous results from GARCH model in that trading imbalance plays a better role than number of trades and trading volume in describing the volatility process. The result is also in agreement with Chan and Fong (2000), although they control order imbalance first in mean regression and then compare the variation of R^2 in the second-stage regression.

The significantly positive coefficients on number of trades, trading volume and the two trading imbalances provide empirical support to the validity of the three information-based hypotheses at daily and intraday intervals. However, when number-based trading variables (Num and I_t^{num}) are applied to the same least square regression, only the coefficients for number of trading imbalance (γ_3) are significantly positive at daily interval. Similarly, when volume-based trading variables ($Volm$ and I_t^{volm}) are in the same regression, only the coefficients for volume of trading imbalance (γ_4) are significantly positive at daily interval. The obtained results here echo the findings based on GARCH model. These comparative regression tests suggest that not only trading imbalance metrics are better explanatory factors to volatility relation but also the asymmetric information hypothesis is more versatile than the other two competing hypotheses[14] [15].

6. Conclusions

The relation between trading activity and volatility has been discussed extensively in the literature. Trading activity has usually been measured by number of trades and trading volume. Some studies have documented a positive relation between price volatility and trading volume and explained the implications based upon information-based hypotheses. In this study, we use different trading variables to examine three prevailing information-based hypotheses, namely, the mixture of distributions, the difference in opinion, and the asymmetric information hypotheses. Because traditional trading variables like number of trades and trading volume can be noisy proxies for information content, and also trading imbalance variables may convey private information from informed trades not revealed directly by number of trades or trading volume, two imbalance metrics are introduced in this paper. Then, our empirical analysis investigates the usefulness of number of trades, trading volume as well as the two trading imbalance metrics. Two sets of experiments are performed with respect to daily and various intraday (ranging from five minutes to one hour) frequencies. The primary experiment, which is based on an array of GARCH models, is used to examine persistent effect and explanatory power of the trading variables. The experiment also checks the validity and, if any, potential dominance of the three information-based hypotheses at different intervals. In addition, since information is random and unpredictable by nature, it is more appropriate to observe the relation between unexpected trading variable and volatility.

Our results conclude that trading imbalance plays a better role than traditional trading variables (i.e., number of trades and trading volume) in explaining volatility-volume relation. Major findings are summarized as follows:

First, significantly positive volume-volatility relations are found by examining number of trades, trading volume and trading imbalances, respectively, in GARCH model and two-stage least square regressions.

---

[14] The coefficients representing existence of the mixture of distributions hypothesis (γ_1) and the difference in opinion hypothesis (γ_2) are positive but not significant at higher frequency, especially at daily interval.

[15] Two-stage least square regressions utilizing trading variables based on total values and expected value are also conducted. Results are similar (but not entirely identical) to the ones from unexpected values, while the explaining powers are lower than the unexpected measures. Hence, it reaffirms that unexpected components may be better proxy for information content than counterparts based on total values and expected values. Results are available on request.

Second, Kurov and Lasser (2004) suggest that, during fast execution of E-mini futures via Globlex system, information contained in the incoming orders is first impounded into the futures contracts. Consistent with their findings, our investigation finds that both trading imbalances exhibit significantly positive relation at all daily and intraday intervals while number of trades and trading volume are significantly positive at only some intraday intervals. Moreover, trading imbalances drastically reduce the persistent effect in GARCH models and demonstrate higher explanatory power in two-stage least square regressions. Results suggest that more information is contained in trading imbalance metrics than in number of trades or trading volume; thereby, they can capture the volatility process better than the traditional variables.

Third, the study evaluates which hypothesis, if any, is relatively more powerful in explaining volume-volatility relation. At the daily interval, only the asymmetric information hypothesis is supported when we introduce pairs of different trading variables to GARCH and least square regressions. At intraday intervals, not only the asymmetric information hypothesis holds, but also the mixture of distributions and the difference in opinion hypotheses prevail at some intraday intervals. These results are consistent with Kurov and Lasser (2004). They suggest that informed traders often derive their informational advantage from order flows. Therefore, more meaningful (private) information embedded in trading imbalances makes the asymmetric information hypothesis dominant at all different intervals.

Fourth, through comparison of results from different frequency intervals, we find that frequency of interval is a non-trivial factor. Because the market is quiet efficient, information may be disseminated and reacted upon over very short time periods. The shorter interval we observe, the more information reflect on volatility we could find.

To the best of our knowledge, this is the first paper to analyze and compare different volatility-volume hypotheses across a multitude of daily and intraday frequencies. In addition, most former studies focus on volume-volatility relation with traditional trading variables. We hope that our results could stimulate more research to explore imbalance-volatility relation through empirical and theoretical work. It follows that there are a few possible extensions of this analysis for further empirical work. First, it would be interesting to find out whether our obtained results are applicable to other E-mini futures contracts and whether there is a gap between regular futures and E-mini futures. Second[16], the measure of volatility used is also a critical factor in examining volume-volatility relation. In this paper, we use conditional variance in GARCH model and the absolute residual from

[16] We are grateful to an anonymous referee for drawing our attention of this issue.

two-stage least square regression to proxy for volatility. We suggest that other measures of volatility, such as realized intraday volatilities[17] (Garman-Klass and 5-minute estimations), can be used to confirm the findings. At the same time, the type of GARCH process can also be an extension issue in the future[18].

[17] Luu and Martens (2003) find that using realized volatility based on intraday returns is more precise than those constructed using daily return. Chen, Daigler and Parhizgari (2006) analyze how three conceptually different types of volatility (they are classical daily volatility, realized intraday volatility and conditional volatility.) affect volatility persistence relationship. They find that realized intraday measure provides the most promising tools for generating highly persistent series for forecasting and derivations valuation.

[18] Chen, Daigler and Parhizgari (2006) also suggest that GARCH technique is very important. The paper compares the performances of GARCH, FIGARCH and FIEGARCH models for financial time series and finds that FIEGARCH model is better than the other two.

References

Arago, V. and L. Nieto, 2005, Heteroskedasticity in the returns of the main world stock exchange indices: volume versus GARCH effects, International Financial Markets Institutions and Money 15, 271-284.

Bessembinder, H., and P. J. Seguin, 1992, Futures-trading activity and stock price volatility, Journal of Finance 47, 2015-2034.

Bessembinder, H., and P. J. Seguin, 1993, Price volatility, trading volume and market depth: Evidence from futures markets, Journal of Financial and Quantitative Analysis 28, 21-39.

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

Chordia, T., R. Roll and A. Subrahmanyam, 2002, Order imbalance, Liquidity and Market Returns, Journal of Financial Economics 65, 111-130.

Chordia, T.,A. Sarkar and A. Subrahmanyam, 2005, An Empirical Analysis of Stock and Bond Market Liquidity, Review of Financial Studies 18, 85-129.

Clark, P. K., 1973, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41, 135-155.

Daigler, R. T. and M. K. Wiley, 1999, The impact of trader type on the futures volatility-volume relation, Journal of Finance 54, 2297-2316.

Epps, T. W. and M. L. Epps, 1976, the stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis, Econometrica 44, 305-321.

Foster, F. D. and S. Viswanathan, 1993, Variations in trading volume, return volatility and trading costs: Evidence on recent price formation models, Journal of Finance 48, 187-211.

Girma, P. B., and M. Mougoue, 2002, An Empirical examination of the relation between futures spreads volatility, volume, and open interest, Journal of Futures Markets 22, 1083-1102.

Grundy, Bruce D., and Maureen McNichols, 1989, Trade and revelation of information through prices and direct disclosure, Review of financial Studies 2, 495-526.

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. Riviv, 1993, Differences of opinion make a horse race, Review of Financial Studies 6, 473-506.

Hasbrouck, J., 2003, Intraday price formation in U.S. equity index markets, Journal of finance 58, 2375-2399.

Holden, C. W. and A. Subrahmanyam, 1992, Long-lived private information and imperfect competition, Journal of Finance 47, 247-270.

Holthausen, Robert W., and Robert E. Verrecchia, 1990, The effects of informedness and consensus on price and volume behavior, Accounting Review 65, 191-208.

Jain, P., Joh, G. 1988, The dependence between hourly prices and trading volume. Journal of Financial and Quantitative Analysis 22, 109-126.

Jones, C., G. Kaul and M. Lipson, 1994, Transactions, volume and volatility, Review of Financial Studies 7, 631-652.

Kalev, P. S., W. M. Liu, P. K. Pham and E. Jarnecic, 2004, Public information arrival and volatility of intraday stock returns, Journal of Banking and Finance 28, 1441-1467.

Kim, Oliver, and Robert E. 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.

Kurov, A. and D. J. Lasser, 2004, Price dynamics in the Regular and E-mini futures markets, Journal of Financial and Quantitative Analysis 39, 365-384.

Lamoureax, C. G. and W. D. Lastrapes, 1990, Heteroskedasticity in stock return data: volume versus GARCH effects, Journal of Business and Economic Statistics 2, 253-260.

Lee, B. and O. M. Rui, 2002, The dynamic relationship between stock returns and trading volume: Domestic and cross-country evidence, Journal of Banking and Finance 26, 51-78.

Luu, J. C, and M. Martens, 2003, Testing the Mixture-of Distributions Hypothesis using "Realized" volatility, Journal of Futures Markets 23, 661-679.

Madhavan, A., M. Richardson and M. Roomans, 1997, Why do security price change? A transaction-level analysis of NYSE stocks, Review of Financial Studies 10, 1035-1064.

Schwet, W., 1990, Stock market volatility and the Crash of 87, Review of Financial Studies 3, 77-102.

Speight, A. E. H., D. G. Mcmillan and O. A. Gwilym, 2000, Intra-day volatility components in FTSE-100 Stock Index Futures, Journal of Futures Markets 20, 425-444.

Tauchen, G. and M. Pitt, 1983, The price variability-volume relationship on speculative markets, Econometrica 51, 485-505.

Watanabe, T., 2001, Price volatility, trading volume, and market depth: evidence from the Japanese stock index futures market, Applied Financial Economics 11, 651-658.

Wood, J. J., McInish, T.H., Ord, J.K., 1985, An investigation of transaction data for NYSE stocks, Journal of Finance 40, 723-739.

Wu, C. and X. E. Xu, 2000, Return volatility, trading imbalance and the information content of volume, Review of Quantitative Finance and Accounting 14, 131-153.

Table 2 Descriptive statistics with respect to different interval frequencies

The data are based on e-mini S&P 500 index futures contracts traded at CME. The data set runs from April 2, 1998 to March 9, 2005. We aggregate data to form observations of daily, hourly, 30-minute, 15-minute and 5-minute frequencies. The respective numbers of observations are noted in the table. R_t is computed from the change in the log of futures price at time t dividend by the log of futures price at time (t-1), where the e-mini S&P 500 index futures price has been adjusted to continued futures contract price. $|\varepsilon_t|$ is volatility measured by two-stage least square regression. All trading variables presented in the table have been de-trended, and the values showing in the table are unexpected component. The unexpected values are residuals from multivariate regression model. Num_t is unexpected number of trades at time t. $Volm_t$ is unexpected trading volume at time t. I_t^{num} is unexpected number of trading imbalance. I_t^{volm} is unexpected volume of trading imbalance.

Variable	Mean	Std. Dev.	Skewness	Kurtosis
Panel A: 5-minute frequency (140714 observations)				
R_t	6.01E-07	0.0014	-0.317	84.130
$\lvert\varepsilon_t\rvert$	0.000903	0.0011	7.794	204.831
Num_t	-2.08E-07	0.1931	2.559	24.422
$Volm_t$	-2.36E-06	3.1005	3.117	88.830
I_t^{num}	-1.81E-07	0.0657	2.514	16.647
I_t^{volm}	-0.003019	1.1238	5.596	109.213
Panel B: 15-minute frequency (46978 observations)				
R_t	1.77E-06	0.0023	-0.4482	31.904
$\lvert\varepsilon_t\rvert$	0.0015	0.0018	5.1794	74.582
Num_t	-2.67E-07	0.4947	1.7301	18.687
$Volm_t$	-2.40E-05	7.9183	1.9481	94.486
I_t^{num}	5.48E-07	0.1107	2.1507	13.441
I_t^{volm}	0.008141	2.0759	4.2320	52.3631
Panel C: 30-minute frequency (24369 observations)				
R_t	3.12E-06	0.0032	-0.3507	21.887
$\lvert\varepsilon_t\rvert$	0.002172	0.0024	4.2443	50.788
Num_t	-5.99E-06	1.0081	1.1284	10.793
$Volm_t$	-9.47E-05	15.539	1.9485	55.587
I_t^{num}	2.64E-06	0.1509	2.1461	13.491
I_t^{volm}	0.013880	3.0971	4.1804	45.052

Table 2 (continued)

Panel D: hourly frequency (13966 observations)				
R_t	5.09E-06 0	.0043	-0.2748	14.815
$\lvert\varepsilon_t\rvert$	0.002923 0	.0031	3.4730	31.851
Num_t	-0.000559 2	.8533	0.8467	5.833
$Volm_t$	-0.000146 25	.5849	1.4082	60.529
I_t^{num}	6.03E-07 0	.2019	2.1922	13.030
I_t^{volm}	0.060378 4	.3144	4.3370	51.569
Panel E: daily frequency (1760 observations)				
R_t	4.55E-05 0	.0128	-0.0815	5.553
$\lvert\varepsilon_t\rvert$	0.009389 0	.0086	1.9728	9.572
Num_t	-0.000681 9	.7659	-0.1184	14.244
$Volm_t$	-0.000212 1	.5279	1.0610	84.637
I_t^{num}	0.000156 0	.5928	1.5508	7.764
I_t^{volm}	-0.002042 14	.2252	1.7049	11.227

Table 3 Correlations for volatility and various trading variables at different interval frequencies

$|\varepsilon_t|$ is volatility measured by two-stage least square regression. Number of trades is the sum of up-tick and down-tick numbers. Trading volume is the sum of up-tick and down-tick volumes. All trading variables presented in the table have been de-trended, and the values presented in the table are unexpected component. Num_t is unexpected number of trades at time t. $Volm_t$ is unexpected trading volume at time t. Number of trading imbalance is the absolute value of the up-tick minus the down-tick numbers. I_t^{num} is unexpected number of order imbalance. Volume of trading imbalance is the absolute value of the up-tick minus the down-tick volumes. I_t^{volm} is unexpected volume of trading imbalance.

	$\|\varepsilon_t\|$	Num	$Volm$	I^{num}	I^{volm}
Panel A: 5-minute frequency					
Num	0.262	1			
$Volm$	0.176	0.734	1		
I^{num}	0.308	0.539	0.320	1	
I^{volm}	0.220	0.511	0.573	0.558	1
Panel B: 15-minute frequency					
Num	0.245	1			
$Volm$	0.178	0.720	1		
I^{num}	0.303	0.395	0.244	1	
I^{volm}	0.230	0.441	0.523	0.508	1
Panel C: 30-minute frequency					
Num	0.222	1			
$Volm$	0.172	0.680	1		
I^{num}	0.307	0.372	0.240	1	
I^{volm}	0.240	0.374	0.487	0.484	1
Panel D: hourly frequency					
Num	0.115	1			
$Volm$	0.164	-0.019	1		
I^{num}	0.303	0.162	0.216	1	
I^{volm}	0.234	0.050	0.464	0.460	1

	$\lvert\varepsilon_t\rvert$	*Num*	*Volm*	I^{num}	I^{volm}
Panel E: daily frequency					
Num	0.099	1			
Volm	0.106	0.605	1		
I^{num}	0.247	0.324	0.213	1	
I^{volm}	0.225	0.177	0.358	0.398	1

Table 3 (continued)

Table 4 GARCH persistence effect on explaining volume-volatility relation (in terms of unexpected components)

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma_1 Num_t + \gamma_2 Volm_t + \gamma_3 I_t^{num} + \gamma_4 I_t^{volm}$$

Frequency	α	β	$(\alpha+\beta)$	γ_1	γ_2	γ_3	γ_4
	0.22	0.80	1.02				
	(288.31)	(1526.5)					
0.19		0.81	1.00	5.06E-07			
	(286.75)	(1526.5)		(224.01)			
5-minute	0.18	0.83	1.01		1.51E-08		
	(291.1)	(1795.9)			(244.67)		
0.16		0.60	0.76			8.45E-06	
	(91.97)	(525.94)				(93.06)	
0.21		0.78	0.99				7.17E-08
	(12.63)	(55.66)					(4.93)
	0.31	0.66	0.97				
	(85.84)	(230.26)					
0.18		0.77	0.95	1.05E-06			
	(95.27)	(470.4)		(147.81)			
15-minute	0.04	0.92	0.96		3.88E-08		
	(99.76)	(314.4)			(122.73)		
0.34		0.58	0.92			4.41E-06	
	(89.37)	(205.57)				(141.61)	
0.17		0.61	0.78				7.74E-07
	(811.1)	(160.53)					(97.45)
	0.01	0.98	0.99				
	(41.41)	(2145.8)					
0.1	1	0.86	0.97	6.45E-07			
	(55.85)	(1037.6)		(60.88)			
30-minute	0.10	0.86	0.96		4.65E-08		
	(54.88)	(728.1)			(130.91)		
0.31		0.42	0.73			1.09E-05	
	(41.56)	(75.94)				(171.29)	
0.22		0.67	0.89				4.51E-07
	(47.08)	(228.7)					(177.06)
	0.02	0.97	0.99				
	(26.55)	(938.9)					
0.19		0.59	0.78	1.28E-06			
	(23.51)	(61.96)		(46.94)			
Hourly	0.03	0.94	0.97		2.74E-08		
	(27.53)	(580.0)			(97.68)		
0.14		0.59	0.73			3.00E-05	
	(11.40)	(25.86)				(17.52)	
0.14		0.59	0.73				9.27E-07
	(23.43)	(153.8)					(69.78)

	0.07	0.91	0.98				
	(8.12)	(86.33)					
0.09		0.90	0.99	6.20E-07			
	(8.63)	(78.58)		(5.02)			
Daily	0.09	0.89	0.98		6.01E-06		
	(8.40)	(73.24)			(6.11)		
0.09		0.89	0.98			1.12E-05	
	(8.29)	(70.36)				(4.75)	
0.09		0.88	0.97				8.38E-07
	(8.33)	(74.37)					(7.17)

Table 5 Comparison of mixture of distributions and asymmetric information hypotheses (unexpected component) using GARCH models

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_1 Num_t + \gamma_3 I_t^{num}$$

Frequency	α	β	$(\alpha+\beta)$	γ_1	γ_3
5-minute	0.22	0.80	1.02		
	(288.31)	(1526.5)			
	0.24	0.62	0.86	5.27E-07*	1.42E-06*
	(281.65)	(663.33)		(252.4)	(210.19)
15-minute	0.31	0.66	0.97		
	(85.84)	(230.26)			
	0.27	0.68	0.95	7.40E-07*	2.25E-06*
	(87.65)	(265.41)		(128.88)	(82.67)
30-minute	0.01	0.98	0.99		
	(41.41)	(2145.8)			
	0.16	0.58	0.74	-3.50E-07	1.50E-05*
	(25.11)	(54.96)		(-16.39)	(53.85)
Hourly	0.02	0.97	0.99		
	(26.55)	(938.9)			
	0.14	0.59	0.73	1.46E-07	1.72E-05*
	(11.03)	(37.85)		(1.81)	(6.78)
Daily	0.07	0.91	0.98		
	(8.12)	(86.33)			
	0.09	0.89	0.98	2.23E-07	8.68E-06*
	(8.45)	(71.08)		(1.32)	(3.01)

t statistics are reported in parentheses.

* indicates that the estimated coefficient is positively significant.

Table 6 Comparison of difference in opinion and asymmetric information hypotheses (in terms of unexpected component) using GARCH models

$$\sigma_t^2 = \omega + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 + \gamma_2 Volm_t + \gamma_4 I_t^{volm}$$

Frequency	α	β	$(\alpha+\beta)$	γ_2	γ_4
5-minute	0.22	0.80	1.02		
	(288.31)	(1526.5)			
	0.16	0.60	0.76	1.28E-08*	1.34E-07*
	(189.21)	(331.19)		(63.58)	(205.20)
15-minute	0.31	0.66	0.97		
	(85.84)	(230.26)			
	0.19	0.60	0.79	3.55E-09*	4.37E-07*
	(78.33)	(310.30)		(7.77)	(315.51)
30-minute	0.01	0.98	0.99		
	(40.86)	(1964.0)			
	0.22	0.61	0.83	2.53E-08*	4.84E-07*
	(49.41)	(222.80)		(41.96)	(152.58)
Hourly	0.02	0.97	0.99		
	(23.81)	(1117.7)			
	0.15	0.59	0.74	-6.44 E-11	9.33E-07*
	(21.13)	(61.59)		(-0.04)	(75.23)
Daily	0.07	0.91	0.98		
	(8.12)	(86.33)			
	0.15	0.57	0.72	-2.27 E-06	2.13E-06*
	(8.06)	(15.87)		(-3.42)	(27.69)

t statistics are reported in parentheses.

* indicates that the estimated coefficient is positively significant.

Table 7 Robustness check on volatility relations (in terms of unexpected components) based on two-stage least square estimation

For daily interval, the dependent variable is the absolute residual from a regression of return on its own 12 lags and day-of-week dummies. The independent variables are 12 lags of the absolute residual, Monday, Tuesday, Wednesday and Thursday dummies, and trading variables. We decompose trading variables into expected and unexpected components, and estimate separately the expected trades variables and unexpected trades variables with volatility. For intraday intervals, the dependent variable is the absolute residuals from a regression of return on its own 12 lags and open and close dummies. The independent variables are 12 lags of the absolute residual, opened and closed dummies, and trading variables. The data cover the period from April 2, 1998 to March 9, 2005. For the sake of brevity, coefficients for the 12 lags of absolute residual and dummies are not reported below. All t-stats are reported in parentheses.

	Frequency									
	5-minute		**15-minute**		**30-minute**		**Hourly**		**Daily**	
Trading variables	**Coeff**	**R^2**	**Coeff**	**R^2**	**Coeff**	**R^2**	**Coeff**	**R^2**	**Coeff**	**R^2**
Num (γ_1)	0.001504		0.000897		0.000524		0.000441		8.59E-05	
	(119.05)	0.31	(62.21)	0.27	(38.83)	0.23	(28.67)	0.20	(4.358)	0.14
$Volm$ (γ_2)	6.30E-05		4.07E-05		2.60E-05		2.04E-05		0.000584	
	(77.93)	0.27	(44.34)	0.24	(29.32)	0.21	(21.26)	0.18	(4.638)	0.14
I^{num} (γ_3)	0.005160		0.004953		0.004889		0.004790		0.003568	
	(141.58)	0.34	(78.71)	0.30	(55.93)	0.28	(41.11)	0.24	(11.33)	0.19
I^{volm} (γ_4)	0.000214		0.000202		0.000188		0.000179		0.000136	
	(97.02)	0.29	(58.68)	0.26	(43.06)	0.24	(32.14)	0.21	(10.25)	0.18

$Num(\gamma_1)$	0.00078 (53.95)		0.000543 (35.88)		0.000292 (20.88)		0.000265 (16.85)		1.67E-05 (0.83)	
$I^{num}(\gamma_3)$	0.0039 (91.31)	0.35	0.0039 9 (59.04)	0.32	0.004164 (44.60)	0.29	0.004098 (33.46)	0.26	0.0035 (10.52)	0.19
$Volm(\gamma_2)$	2.76E-05 (28.31)	1	.78E-05 (16.84)		1.02E-05 (10.26)		8.09E-06 (7.63)	0.000	149 (1.13)	
$I^{volm}(\gamma_4)$	0.00017 (63.42)	0.29	0.00016 7 (41.33)	0.26	0. 000163 (32.71)	0.25	0.000157 (24.96)	0.21	0.000131 (9.24)	0.18

Appendix

Mixture of distributions hypothesis

The conceptual framework was developed and explored by Clark (1973), Epps and Epps (1976), and Tauchen and Pitts (1983). Basically, the models of Clark (1973), Harris (1987), and Tauchen and Pitt (1983) link the number of trades to the number of information events. Harris (1987) extends further this framework and introduces multiple mixing variables to the relation. He finds that volume per transaction and number of transactions are useful factors in modeling. Later, Jones, Kaul, and Lipson (1994) show that frequency of trades can determine volatility of returns. They also discover that size of trades conveys virtually no information beyond what has been contained in number of trades. Essentially, their findings support the mixture of distributions hypothesis. Similarly, the results of Chan and Fong (2000) also support the existence of the hypothesis and conclude that number of trades in an influential factor in the volume-volatility relation. The foundation of the hypothesis is based on the assumption of a joint bivariate normal distribution for volume and volatility conditional upon the arrival of information, which induces changes in volume and volatility accordingly. Innovation to this volume-volatility relation is caused by a mixing variable, usually measured by the number of information arrivals. As a summary, the mixture of distributions hypothesis can be confirmed by a significantly positive relation between number of trades and volatility.

Difference in opinion hypothesis

The framework is based on the notion that investors possess different opinions and interpretations of information. Harris and Raviv (1993) show that a greater dispersion of beliefs creates price variability and volume excessive to the equilibrium level. Since traders hold different opinions on information, trading takes place when public information switches from one state to another (for example, favorable to unfavorable condition, or vice versa.) Therefore, trading volume and volatility are positively related because both are correlated with the arrival of public information. In other words, the difference in opinion models can be verified by a positive relation between trading volume and volatility. Kim and Verreccia (1991), Holthausen and Verrecchia (1990), and Grundy and Mcnichols (1989) are some other representative work in this stream of study.

Asymmetric information hypothesis

This group of models considers trading as a result of asymmetric information within the market. The basic premises are that factors such as size of trade and order imbalance convey information about the degree of information asymmetry, and that this asymmetry cannot be directly obtained / deduced from trading volume or number of trades. Within this framework, there are informed and uninformed market traders in an asymmetric information environment. Informed traders have relatively homogeneous beliefs, of which they base their knowledge on the market and fundamental characteristics of the assets. If more informed traders are confident of the information they possess, their orders will cluster at one side of trading. Such order imbalance will induce a drastic change in asset prices. Thus, order imbalance should reflect the quality of private information, and hence, affect return volatility. Holden and Subrahmanyam (1992) show a positive relation between size of trades and quality of information possessed by informed traders. Since size of trades and order imbalance are likely to be positively related to the quality of information, the variables are correlated with price volatility in the same manner. Wu and Xu (2000) introduce transaction and volume imbalances to capture return volatility. They show that trading imbalances have strong explanatory power on modeling return volatility. They also suggest that trading imbalances contain non-trivial information about private signals. In short, trading imbalance variables can be used to examine the volatility process and thus the asymmetric information hypothesis. In a similar study, Chan and Fong (2000) find that size of trade and order imbalance play significant roles in the volume-volatility relation. Their study also provides support to the asymmetric information model in the equity market.