Public News Arrival and Cross-Asset Correlation Breakdown

Kin-Yip Ho, Wai-Man Liu, and Jing Yu*

This Version: April 2014

*Ho and Liu are from the Research School of Finance, Actuarial Studies and Applied Statistics, CBE Building 026C, The Australian National University (ANU), Canberra ACT 0200, Australia. Yu is from the Discipline of Accounting and Finance, The University of Western Australia Business School, 35 Stirling Highway, Crawley, WA 6009, Australia. Ho: kin-yip.ho@anu.edu.au, tel: +61-2-61257299; Liu: waiman.liu@anu.edu.au, tel: +61-2-61253471 Yu: jing.yu@uwa.edu.au; tel: +61-8-64881759. The authors would like to thank Jonathan Brogaard, Kyoung-Shik Eom, Carole Comerton-Forde, Doug Foster, Bruce Grundy, Richard Heaney, Terrence Hendershott, Philip Maymin, Phong Ngo, Geoff Warren, Jaeho Yoon, Yinggang Zhou, and conference participants at the 10th International Paris Finance Meeting, 7th Annual Conference on Asia-Pacific Financial Markets of the Korean Securities Association, and seminar participants at ANU, Chinese University of Hong Kong, and Ewha Woman's University. We would like to thank Yanlin Shi from ANU for excellent research assistance.

Public News Arrival and Cross-Asset Correlation Breakdown

Abstract

This study models and tests empirically the role of public news arrivals in the quote matching across single stock futures and underlying stock markets - a trading strategy often adopted by algorithmic traders. Our model suggests that quote return correlation across these two markets breaks down when the news uncertainty is sufficiently large and futures market makers switch from automating the quote matching process to manually analyze, monitor and update quotes. We show empirically that the breakdown is more prominent for large stocks, and this effect of firm size falls during periods of high market volatility.

Keywords: Correlation Breakdown, Algorithmic Quote Matching, Single-Stock Futures, Public News Arrival

JEL Classification Number: G10, G12, G14

1. Introduction

High speed of trading and an increasing pace of information dissemination in the marketplace are two most significant developments of our financial environment largely due to the rapid rise of algorithmic trading.[1] As a new form of trading scheme, algorithmic trading utilizes computer-based algorithms to implement high frequency trading strategies without human intervention. A young but growing literature emerges to understand the market impact of high frequency trading environment. Despite their mixed conclusions, an important assumption underlying the theoretical work in this strand of literature is that algorithmic trading can respond promptly to market news events and process news faster than their human counterparts (e.g., Biais, Foucault and Moinas (2011), Foucault, Kadan, and Kandel (2013), and Jovanovic and Menkveld (2012)).

However, the market reactions to the erroneous report on the United Airlines (UAL) prompt us to pause and rethink about this assumption: on September 8 2008, a headline announcing the bankruptcy of the airline hit the news feeds by mistake. Within a short span of 12 minutes, the UAL share price plummeted by 75% to $3 a share before it subsequently recovered. As suggested by the *New York Times* on September 13 2008, "human error seems to have played only a minor role. The financial damage was mostly the result of the interplay between the algorithms that search and compile information from the Web and

[1] By 2009, algorithmic trading has dominated a substantial portion of equity trades in the U.S., with more than 70% of all equity trades being generated by orders from computer models reacting in sub-seconds to market news and real-time price movements (Donefer (2010), Hendershott, Jones and Menkveld (2011)). These strategies are estimated to represent industry revenues of around US$7-$9 billion. For the first quarter of 2010, at least one third of the order book executions on the London Stock Exchange (LSE) are resulted from algorithmic trading. The market share of algorithmic trading on LSE is obtained from the exchange's responses to Committee of European Securities Regulators' Call for evidence on micro-structural issues of the European equity markets (April 30 2010). Website: http://www.londonstockexchange.com/about-the-exchange/regulatory/responsetocesrscallforevidenceonmicro-structuralissuesoftheeuropeanequitymarkets.pdf (Viewed July 4 2013). Similar fast growing pattern of algorithmic trading is also seen in foreign exchange and financial derivatives markets (Chaboud, Chiquoine, Hjalmarsson, and Vega (2013), Gomber, Arndt, Lutat, and Uhle (2011)).

the ones that Wall Street firms and hedge funds use to make trades automatically". This story challenges our conventional view of how algorithmic trading handles information flows in reality. The immediate questions are: can algorithmic trading conducted by machines react appropriately at the arrivals of complex public news which require advanced analytical interpretation? Under what circumstances are these news arrivals relevant?

Our study attempts to tackle these questions by modeling and testing empirically the role of public news arrivals in a trading environment populated by algorithmic trading. Broadly speaking, there are two types of algorithmic trading. One type of algorithmic traders uses computers to reduce transactions costs (e.g., Cvitanic and Kirilenko (2010)) while the other plays the role of arbitragers who constantly seek for mispricing opportunities (e.g., Jarrow and Protter (2012)). Our paper relates to the latter type. Specifically, we are interested in studying the trade-offs that market makers of two closely related assets face when they use algorithmic quote updating to reduce mispricing risk, and factors that might alter the relationship between news arrivals and the return correlations between these two assets. The asset pair used in our empirical setting are the single stock futures (SSF) contracts and their underlying stocks since in normal circumstances their return correlation should be close to perfect and are not subject to time variation (as compared to, e.g., stock options). Hence it gives us a cleaner measure of correlation. Also, studying the SSF market enables us to examine how cross-sectional firm characteristics affect the news-correlation dynamics.

The intuition behind our model is straightforward. By the cost-of-carry relation, futures market makers peg the quotes to the quotes of the underlying automatically using computerized algorithms. However, futures market makers may shift from quote pegging to manual monitoring of the news feed upon the arrival of public news if the news content is vague about the asset's fundamental value. Monitoring the news feed and analyzing its impact on prices can be a costly exercise because the correct interpretation of an announcement requires human attention and processing (see Foucault, Röell, and Sandås (2003) and Liu (2009) for theory; see Chakrabarty and Moulton (2012) for empirical support). Because news monitoring is costly, our model suggests that, futures market makers will widen the spread to compensate for the cost of monitoring, and ultimately cause a momentary breakdown in spot-futures return correlation.[2]

Our theoretical model yields three empirical implications. First, the strong contemporaneous return correlation between spot and futures declines as public news arrives and the dispersion of agreements on public news heightens, provided that the cost of manual monitoring is not too high. Second, the return correlation breakdown is more likely to occur for large stocks which are characterized by lower costs of monitoring (Size Effect). Finally, we show that the Size Effect weakens when the overall market is experiencing turbulence (i.e. when the opportunity costs of not monitoring other stocks are high).

We test these empirical implications using high frequency transactions data of all U.S. SSF contracts listed on One-Chicago Futures Exchange and their underlying stocks from Thomson Reuters Tick History (TRTH) database, and real time public news database from RavenPack News Analytics Dow Jones Edition (RavenPack) database between 20 July 2009 (when the SSF quote data becomes available) and 19 March 2010 (when the March 2010 contract expires). To gauge the realized spot-futures return correlations, we design two conditional correlation measures that are computed with a forward rolling window of six 10-minute time intervals. The first measure is based on the logarithmic returns of spot and futures quotes and the second measure is based on the residuals from filtering the logarithmic returns of spot and futures quotes through the VEC-BEKK model. The latter construct is

[2] One might argue that at the arrival of news, market makers do not shut down its algorithms entirely, but instead use some types of news analytics or other forms of black box algorithms to avoid human intervention. This may be true but often it is necessary for market makers to reset parameters of their algorithms to reflect the news contents, and this requires, albeit reduced, human intervention.

intended to address the cost of carry relation and the cointegrating properties of the spot and futures prices. For each stock, we compute the daily conditional correlation measures by taking an average of intraday rolling window conditional correlations.

Our comprehensive public news data come from the RavenPack database. This database covers a wide spectrum of U.S. stocks and keeps track of news stories that run through major real-time newswires such as Dow Jones Newswires, regional editions of the Wall Street Journal, and Baron's and other internet sources including financial sites, blogs, local and regional newspapers with millisecond time-stamps. Using a proprietary news analytic algorithm, RavenPack develops a set of metrics that quantify the relevance and the sentiment of each news story based on its textual content. In our empirical tests, we design news variables to address two dimensions of public news: the count of public news and the interpretation of news. Specifically, the first set of news proxies account for the number of news events per day whilst the second set of news proxies compute the standard deviation of news sentiments, which provides a proxy for the disagreements of opinions among investors.

In line with our theoretical claim, we document compelling evidence that all measures of spot-futures return correlation significantly decline as the number of firm-specific public news rises. This result remains robust after controlling for a variety of stock characteristics. It is possible that the number of news events might not accurately reflect the uncertainty about the news content if all the news articles cover similar stories. Therefore, to capture the dispersion of beliefs over news events we empirically use another news measure - volatility of news sentiments. Again, we find consistent evidence of spot-futures return correlation breakdown associated with an increase in the disagreement about the interpretation of public information. Moreover, even after controlling for the number of relevant news messages per day the significance of news sentiment volatility variable is not subsumed. This finding is robust to different news sentiment measures used to proxy for news uncertainty.

After establishing the linkage between public news arrivals and spot-futures return correlation breakdown, we turn to examine whether and how certain firm attributes and market circumstances pertaining to the direct cost of monitoring the stock and the opportunity cost of not monitoring other stocks influence the news effect on return correlation breakdown. It is argued that small-cap stocks require more costly monitoring from market makers than large-cap stocks because of their lack of information transparency (Hong, Lim, and Stein (2000), Dechow and Dichev (2002) and Gadarowski (2002)). Further, since attention is a scarce cognitive resource (Kahneman (1973)) and market makers must consider the optimal balance between the direct cost of monitoring and the opportunity cost of not monitoring other stocks. During turbulent times, the attention of market participants is dominated by market-wide or macro-economic news because they face a higher opportunity cost of not monitoring these news which may affect the value of other stocks[3]. A higher opportunity cost implies a lower picking off risk as the distracted arbitrageurs and news watchers are preoccupied with other news events and pay less attention to stock-specific news. As a result, market makers become less sensitive to the effect of monitoring cost on quote updating strategies. Our model predicts that the size effect is weaker when the overall market is experiencing turbulence. In support of these arguments, our empirical evidence shows that at times of significant firm-specific news arrivals, return correlation breakdown appears to be more pronounced for large-cap stocks but the news-correlation sensitivity on large stocks is less prominent in the case of extreme market turbulence.

Our study makes an important contribution to the recent literature in algorithmic trading

[3]During volatile market conditions, market participants tend to shift the attention from individual stocks to the overall market news and performance due to economic or psychological reasons. Increased comovements in stock returns and liquidity during market stress provide valuable implications for this notion. For example, Ang and Chen (2002) document a rise in individual stock return comovements during volatile markets while Hameed, Kang, and Viswanathan (2010) and Karolyi, Lee, and van Dijk (2012) report an increase in commonality in liquidity (i.e., systematic liquidity dries out across individual stocks in and outside the US.)

from both theoretical and empirical perspectives. Prompted by the dominance of algorithmic trading, a growing body of theoretical studies attempt to address the debatable effects of algorithmic traders.[4] An embedded assumption in many studies of this line is the relative speed advantage of algorithmic traders as opposed to human traders. For example, Jovanovic and Menkveld (2012) theorize that algorithmic traders react faster than human traders to market signals and therefore improve the information efficiency in the marketplace. Under the similar assumption, Biais, Foucault and Moinas (2011) model the trading equilibrium with the presence of high frequency traders and conclude that the fast information processing enables these traders to exploit (slow) liquidity traders, thus giving rise to higher adverse selection costs. Foucault, Kadan and Kandel (2013) theorize that fast information gathering by algorithmic traders considerably reduces the monitoring costs of investors and market makers. Specifically focusing on the arbitrage strategies of algorithmic trading like our study, Gerig and Michayluk (2010) lend support to the liquidity-improving effects of algorithmic traders on the condition that algorithmic traders are able to collect information simultaneously at different exchanges and/or in different but related securities.

Our theoretical and empirical findings, however, have reservation on the feasibility of this assumption. We show that cross-asset correlation breaks down when news uncertainty is high. This is a strong indication of withdrawal of algorithmic traders in replacement with human counterparts when news uncertainty is high. Our results are in line with empirical findings in Chaboud, Chiquoine, Hjalarsson, and Vega (2013), which reports a decline in liquidity supply from algorithmic traders in the foreign exchange market immediately after the economic news arrivals.

Our research also contributes to the literature on the asset pricing implications of public news arrival. One critical challenge in the analysis of media content is to transform quali-

[4]See Hasbrouck and Saar (2013) for an excellent review.

tative news contents into quantitative variables suitable for econometric estimations. With technological advancement, a small number of studies emerge to study quantified media content using computer algorithms. For instance, Tetlock (2007) investigates the news sentiment on the daily content of a popular Wall Street Journal column and finds that high media pessimism predicts lower market prices followed by a reversal to fundamentals. Riordan, Storkenmaier, Wagener and Zhang (2013) analyze the tone of news messages and document that negative messages are associated with higher adverse selection costs than positive messages. We are among the first to infer the dispersion of investor opinions from news content. Antweiler and Frank (2004) also examine the disagreements on public information on stock trading but different from ours, they infer the disagreements from internet stock message boards rather than newswire messages. We establish a significant link between the dispersion of investor opinions and algorithmic trading, consistent with finance theory that disagreements among investors about public information induce stock trading activities (Harris and Raviv (1993), Kandel and Pearson (1995), and Cao and Ou-Yang (2009)).

The remainder of the paper is structured as follows. Section 2 presents a simple model of futures-spot dynamics and news monitoring. Section 3 describes our sample, data sources and variable construction. Section 4 discusses the empirical results on the link between public news arrival and correlation breakdown, and section 5 concludes.

2. A Simple Model

2.1. Value process and model assumptions

Consider a risky asset that trades in two markets: the asset is traded in the spot market, while a futures contract is written underlying the spot. For simplicity, the level of interest rate is assumed to be zero and the stock does not pay any dividend. Denote as T the maturity of the contract. The trading interval $[0, T]$ of both spot and the futures is partitioned into T equally spaced subintervals: $[0, 1]$, $[1, 2]$, ..., $[t, t+1]$, ..., $[T-1, T]$. At the beginning of each subinterval, a piece of public news is released but its impact on the asset's value is uncertain. In this game, there are two competitive market makers. One is for the spot and the other is for the futures, and they operate and trade separately in their respective markets. Market makers may monitor the news flow, and adjust the quotes before the news is being studied. Meanwhile, traders who are informed of the news may arrive and compete to pick off the quotes before its adjustment. The game then ends and the uncertainty about the news is resolved. The game will be repeated in the next subinterval $[t+1, t+2]$, and continues until time T.[5]

Denote as V_t the value of the asset prior to the news release at the subinterval $[t, t+1]$. Subsequent to the news arrival, the value of the asset follows a trinominal process:

$$V_{t+1} = \begin{cases} V_t + \frac{\sigma_t}{2} \\ V_t \\ V_t - \frac{\sigma_t}{2} \end{cases} \quad \text{with probabilities:} \quad \begin{matrix} \frac{\alpha}{2} \\ 1-\alpha, \\ \frac{\alpha}{2} \end{matrix} \tag{1}$$

where $\sigma_t > 0$ and $\alpha \in (0, 1)$. σ_t captures the uncertainty about the impact of the news arriving in the interval $[t, t+1]$. The equilibrium spread is determined depending on the monitoring strategy of the spot and futures market makers.

2.2. News monitoring and quote updating strategies

Subsequent to the arrival of news, two strategies of monitoring are considered. Under the first strategy (Strategy 1), risk-neutral market makers of spot and futures monitor the news

[5] In this game, we do not assume the spot and the futures market share the same market maker because such equilibrium is unstable. To see why, let us assume the futures market is less liquid than the underlying spot and quotes in both markets are set by the same market maker, and she monitors for the news flow in setting both quotes. Because of the differences in liquidity, the spread of the futures contract must be wider. However, as shown in (9), her futures quotes are susceptible to be undercut by other competitive market makers who use algorithmic quote updating. This outcome implies that in equilibrium, we would have one market that comprises market makers conducting news monitoring, while another market that comprises market makers conducting algorithmic quote matching.

separately. They do not learn from each other. The trading/monitoring game is divided in two stages. At stage 1 (quotation stage), market makers set quotes. At stage 2, they exert some efforts into monitoring to evaluate the impact of the arriving news. Immediately after the news release, competitive market makers set quotes based on zero expected profit condition. Subsequent to the quotation stage, market makers must exert some monitoring efforts into studying the nature of the news release. The level of monitoring efforts affects the probability that the trader understands fully the news impact and reacts before others. We label this probability λ, for $\lambda \in (0, 1)$. Nature reveals the news impact to market makers according to their chosen probabilities. Monitoring activity requires traders to incur an opportunity cost, which can be interpreted as the monetary value associated with the level of disutility caused by the efforts exerted to monitor the news release. We label this cost $K(\lambda)$, for $K'(\lambda) > 0$ and $K''(\lambda) > 0$. To make the model tractable, we consider a quadratic monitoring cost function:

$$K(\lambda) = \frac{c\lambda^2}{2}. \tag{2}$$

The monitoring cost function $K(\lambda)$ consists of the research efforts exerted in learning the "informativeness" or the "quality" of the news observed. For instance, traders may need to spend time to understand and filter the noisy signal inherent in accounting numbers when an earnings report is released.

Under the second strategy (Strategy 2), the spot market maker monitors the news flow, while the futures market maker invests a technology that enables her to match the quotes (almost) instantaneously against the quotes of the underlying spot. While this may create a disincentive to the spot market maker to monitor, the technology is costly and is not readily available for everyone. In the model, we assume the futures market maker possesses this technology. The investment cost is sunk. The marginal cost is fixed per trade but it is negligibly small. This assumption is reasonable in the setting of SSF traded on One-Chicago Futures Exchange. In One-Chicago, there are only three lead SSF market makers who are responsible for supplying liquidity on futures contracts that are written on more than two thousand stocks.[6] Since it is beyond humanly possible for them to manually monitor the flow of firm-specific news of more than 2,000 stocks, they have to rely on automated quote updating using some kind of algorithms. In contrast, the spot market comprises many market makers who could potentially specialize in tracking news stories of different stocks, making news monitoring less impossible.[7]

2.3. Equilibrium

Next, we derive the equilibria associated with these two strategies. Without loss of generality, we assume that the default strategy is Strategy 2, i.e. algorithmic quote updating. We show that there will be a switch in strategy (from Strategy 2 to Strategy 1) when news uncertainty rises beyond a threshold. The futures market maker will "switch off" the algorithm, widen the bid-ask spread and study the news. Consequently, the correlation between the spot and the futures quote return drops. When news uncertainty falls, the futures market maker "switches" the algorithm back. The spread then narrows and the correlation increases.

2.3.1. Equilibrium under Strategy 1

We assume there exists four risk-neutral liquidity providers (liquidity buyer and seller in futures and spot markets). At the beginning of the interval $[t, t+1]$, quotes are determined by

[6] They are the Chicago Trading Company (CTC), Quiet Light Securities and Timber Hill. See: http://www.onechicago.com/?page_id=53 (Viewed July 4 2013).

[7] In this paper, we do not model the strategy which involves matching the spot quotes against the underlying futures. While it is possible theoretically, the futures tend to follow the spots for the majority of SSF contracts. That said, the interpretation of our model can easily be modified if the liquidity of the futures market is higher than that of the spot market.

setting the expected payoff for spot liquidity seller ($\Pi_{a,s}$), spot liquidity buyer ($\Pi_{b,s}$), futures liquidity seller ($\Pi_{a,f}$) and the futures liquidity buyer ($\Pi_{b,f}$) to zero. Define A_t^j and B_t^j as the best ask and the best bid prices for contract j at the beginning of the interval, for $j = s, f$. Define S_t^j as the bid-ask spread for contract j, and due to symmetry, $\frac{S_t^j}{2} = V_t - B_t^j = A_t^j - V_t$. Therefore, zero expected payoff condition implies:

$$\Pi_{i,j} = -\frac{\alpha}{2}(1 - \theta\lambda_{i,j})\frac{1}{2}(\sigma_t - S_t^j) + (1 - \alpha)\frac{\beta_j}{2}S_t^j - K(\lambda_{i,j}) = 0, \forall i = a, b \text{ and } \forall j = s, f. \quad (3)$$

β_s (β_f) is the probability of liquidity traders arriving at the spot (futures) market when no news arrives. The first term in the expected payoff function (3) represents the expected cost of being picked off. This cost is also known as the adverse execution cost when bad (good) news arrives. When the prevailing bid (ask) quote is too high (low), traders who learn of the news will "pick off" the mispriced bid (ask) quote. The second term represents the expected gain from providing liquidity. The last term represents the cost of monitoring.

In the payoff function (3), θ relates to the probability that quotes are picked off when news monitoring is perfect, i.e. $\lambda = 1$. θ is low when the quote updating process is slow, or when competition among arbitragers is intense. In both cases, the probability of being picked off is high. θ also relates to *the opportunity cost of not monitoring other stocks*. When the overall market is experiencing extreme movements, traders may shift their focus from monitoring specific stocks to events that cause the market-wide volatility. In this case, θ is expected to be large. Conversely, when there are events that shift traders' focus to the stock because the opportunity cost of not monitoring other stocks is low, we expect more intense competition among arbitragers on that stock and thus θ is low.

In the second stage, market makers must solve for the optimal λ given the quotes set in the first stage:

$$\max_{\lambda_{i,j} \in (0,1)} \frac{\alpha}{4}\theta\lambda_{i,j}(\sigma_t - S_t^j) - \frac{c\lambda_{i,j}^2}{2}, \forall i = a, b \text{ and } \forall j = s, f. \quad (4)$$

and the interior solution to the optimization problem (4) is:

$$\lambda_{i,j} = \frac{\theta\alpha}{4c}(\sigma_t - S_t^j) \tag{5}$$

Substituting the optimal λ into the expected payoff function, the zero expected profit condition (3) yields the optimal spread for the spot and the futures under Strategy 1, $S_t^j(1)$:

$$S_t^j(1) = \sigma_t - \frac{8c}{\theta^2\alpha^2}\left\{(1-\alpha)\beta_j + \frac{\alpha}{2} \pm \sqrt{\left[(1-\alpha)\beta_j + \frac{\alpha}{2}\right]^2 - \frac{\theta^2\alpha^2(1-\alpha)\beta_j\sigma_t}{4c}}\right\}, \forall j = s, f. \tag{6}$$

Note that we can ignore one of the quadratic roots because the spread is bounded above by σ_t, as σ_t represents the maximum cost the market maker faces when her quotes are picked off for certain.When α approaches one, $S_t^j(1)$ approaches σ_t. When σ_t approaches zero, $S_t^j(1)$ approaches zero. Therefore the equilibrium spread $S_t^j(1)$ must be given by:

$$S_t^j(1) = \sigma_t - \frac{8c}{\theta^2\alpha^2}\left\{(1-\alpha)\beta_j + \frac{\alpha}{2} - \sqrt{\left[(1-\alpha)\beta_j + \frac{\alpha}{2}\right]^2 - \frac{\theta^2\alpha^2(1-\alpha)\beta_j\sigma_t}{4c}}\right\}, \forall j = s, f. \tag{7}$$

The equilibrium quotes for spot and futures at the beginning of the interval (just after the news release but before the impact of the news was studied by the market maker) under Strategy 1 are:

$$A_t^j(1) = V_t + \frac{1}{2}S_t^j(1) \text{ and } B_t^j(1) = V_t - \frac{1}{2}S_t^j(1), \forall j = s, f. \tag{8}$$

2.3.2. Equilibrium under Strategy 2

Suppose the futures market maker can automate the quote updating process without proceeding to manual monitoring. For the futures-spot no-arbitrage condition to hold, futures market maker must create a quote updating algorithm to hedge against adverse movements in the underlying spot prices. In doing so, she must incur a fixed cost of maintaining a system that automates the quote updating algorithm. While the fixed cost is large, it is sunk and we assume the marginal cost is zero. Define $A_t^f(2)$ and $B_t^f(2)$ as the futures ask and bid quotes under Strategy 2. They are:

$$A_t^f(2) \geq A_t^s(1) \text{ and } B_t^f(2) \leq B_t^s(1) \tag{9}$$

Note that in equilibrium, the equality is always binding. If $A_t^f(2) < A_t^s(1)$, spot market maker can always make non-zero profits by buying at $A_t^f(2)$ with a market order and selling at the spot's quote: $A_t^s(1)$. Meanwhile, zero expected profit condition implies that competition will prevent $A_t^f(2)$ from exceeding $A_t^s(1)$ since futures market maker will have an incentive to undercut and drive the ask quote down to parity.

$A_t^f(2)$ faces a higher risk of being picked off. The payoff of algorithmic quote matching depends on the trade-offs between (i) the expected cost of being picked off when the market maker fails to match the spot quotes promptly enough, (ii) the cost of algorithmic quote updating process, and (iii) the gain from providing liquidity. Mathematically:

$$\Pi_{a,f} = -\frac{\alpha}{2}\theta\big[(V+\frac{\sigma_t}{2})-A_t^f(2)\big]-\frac{\alpha}{2}(1-\theta)(A_t^s(1)-A_t^f(2))+(1-\alpha)\frac{\beta_f}{2}(A_t^f(2)-V) \geq 0 \tag{10a}$$

$$\Pi_{b,f} = -\frac{\alpha}{2}\theta\big[B_t^f(2)-(V-\frac{\sigma_t}{2})\big]-\frac{\alpha}{2}(1-\theta)(B_t^f(2)-B_t^s(1))+(1-\alpha)\frac{\beta_f}{2}(V-B_t^f(2)) \geq 0 \tag{10b}$$

The first and the second terms of (10a) and (10b) reflect the expected payoff if the futures quotes are picked off when the value innovation is $+\frac{\sigma_t}{2}$ $(-\frac{\sigma_t}{2})$. The first term reflects the expected payoff when they fail to match with prevailing spot quotes (with probability θ), while the second term reflects the expected payoff when the quotes are matched with that of the cash market (with probability $1-\theta$). If (9) is binding, the second term drops out. The futures market maker could go long (short) in the cash market to hedge against her short (long) futures position.

After a considerable algebra, it can be shown that :

$$\Pi_{a,f} < 0 \text{ and } \Pi_{b,f} < 0 \text{ iff } \sigma_t > \sigma^* \tag{11}$$

where:

$$\sigma^* \equiv 4c\left[\frac{\alpha\theta + 2(1-\alpha)\beta_f}{\theta^2\alpha(1-\alpha)\beta_f}\right]\left(1 - \frac{\beta_s}{\beta_f}\theta\right) \tag{12}$$

and $\sigma^* > 0$ if the following condition holds:

$$1 - \frac{\beta_s}{\beta_f}\theta > 0 \tag{13}$$

(11) is intuitively appealing. If the spot is highly liquid (i.e. β_s is large), the spot spread is very tight, thereby making the payoff of the automatic quote updating strategy more likely to be negative when σ_t rises. If β_f is large however, σ^* is high, the payoff of the strategy is less likely to be negative when uncertainty rises.

Let us assume, at the beginning of the interval $[t, t+1]$, $\sigma_t < \sigma^*$, and thus: $\Pi_{b,f} = 0$ and $\Pi_{a,f} = 0$, Strategy 2 is the equilibrium strategy. If news uncertainty remains below this threshold in the subsequent interval, futures and spot quote returns are expected to be perfectly and positively correlated. Take the ask quotes as an example. At $t+1$, the return of the futures ask quote is the same as the return of the spot ask quote:

$$A^f_{t+1}(2) - A^f_t(2) = A^s_{t+1}(1) - A^s_t(1), \tag{14}$$

Now suppose news uncertainty rises in the interval $[t+1, t+2]$ such that $\sigma_{t+1} > \sigma^*$. The futures market maker will shift from Strategy 2 to Strategy 1 because the expected payoff associated with the algorithmic quote matching system is negative. In this case, the futures ask quote return will be:

$$A^f_{t+2}(1) - A^f_{t+1}(2) = A^s_{t+2}(1) - A^s_{t+1}(1) + \epsilon(\sigma_{t+1}), \tag{15}$$

where $\epsilon(\sigma_{t+1}) \neq 0$ as:

$$\epsilon(\sigma_{t+1}) = \frac{4c}{\theta^2\alpha^2}\left[(1-\alpha)\left(\beta_s - \beta_f\right) + \sqrt{Y_f(\sigma_{t+1})} - \sqrt{Y_s(\sigma_{t+1})}\right], \tag{16}$$

for $Y_j(\sigma_t) \equiv \left[(1-\alpha)\beta_j + \alpha\right]^2 - \frac{\theta^2\alpha^2(1-\alpha)\beta_j\sigma_t}{2c}$, $\forall j = s, f$. Further analysis shows that a significant rise of uncertainty would lead to a positive $\epsilon(\sigma_{t+1})$ as:

$$\epsilon(\sigma_{t+1}) > 0 \text{ iff } \sigma_{t+1} > \frac{8c}{\theta^2\alpha}. \tag{17}$$

A positive ϵ implies that, as the futures market maker shifts the strategy to manual monitoring, futures spread must widen to compensate for the cost of monitoring. We expect a correlation breakdown as returns will no longer be perfectly correlated. The same argument applies for the bid quote returns.

We now establish the first hypothesis.

***Hypothesis* 1**: *(News Effect) As news uncertainty rises, the likelihood of a shift from algorithmic quote matching to manual monitoring in the futures market increases, and resultantly, the contemporaneous correlation between the spot and the futures return is more likely to fall.*

Hypothesis 1 also implies that when news uncertainty falls (below the threshold), Strategy 2 becomes the equilibrium strategy, and using expression (9), the contemporaneous correlation reverts to one as:

$$A^f_{t+1}(2) - A^f_t(2) = A^s_{t+1}(1) - A^s_t(1), \tag{18a}$$

$$B^f_{t+1}(2) - B^f_t(2) = B^s_{t+1}(1) - B^s_t(1). \tag{18b}$$

In addition, we would expect the correlation between the futures return and the spot return falls more for small stocks than for large stocks. This is because small stocks may receive little investor attention (Hong, Lim, and Stein (2000)), have a lower quality of financial reporting (Dechow and Dichev (2002))[8], release less public information, and have a

[8]Hong, Lim, and Stein (2000) suggest that private information travels faster for large stocks due to the fact that investors collecting private information are faced with the fixed costs of information acquisitions.

lower extent of financial press coverage (Gadarowski (2002))[9]. Thus, traders must exert a greater amount of labor efforts in assessing the news impact on the quotes of small stocks. Therefore, we would expect small stocks to have a lower cost in news monitoring (higher c). Differentiating σ^* with respect to c yields:

$$\begin{aligned} \frac{\partial \sigma^*}{\partial c} &= 4\left[\frac{\alpha\theta + 2(1-\alpha)\beta_f}{\theta^2\alpha^2(1-\alpha)\beta_f}\right]\left(1 - \frac{\beta_s}{\beta_f}\theta\right) \\ &> 0. \end{aligned} \tag{19}$$

Since small stocks, on average, have a higher c. $\partial\sigma^*/\partial c > 0$ implies that small stocks would have a smaller chance of a correlation breakdown. For large stocks the threshold value is lower and therefore the likelihood of σ_{t+1} exceeding the threshold value is expected to higher. This leads us to the second hypothesis:

***Hypothesis* 2**: (*Size Effect*) *As news uncertainty rises, the likelihood of a shift from algorithmic quote matching to manual monitoring in the futures market is higher for large stocks. Thus, the contemporaneous correlation between the spot and the futures return is more likely to fall for large stocks.*

Given that correlation breakdown is an infrequent event and the fact that large stocks have a higher probability of a correlation breakdown, the unconditional average news effect on the *level* of contemporaneous correlation is expected to be larger for large stocks.

Lastly, we show that the *Size Effect* varies depending on the opportunity cost of not

As a result, they are more inclined to devote more efforts to stocks they can take large positions. Another strand of the rich literature on financial accounting quality (e.g., Dechow and Dichev (2002)) supports the view that financial accounting quality is relatively high for large stocks.

[9]Analysts collect, filter and disseminate information about firms. They cover large stocks instead of small stocks because there is a high demand of analyst service from institutional investors who trade stocks that are in the index funds. They cover large stocks also because the cost of collecting non-public information is lower for large stocks. The relation between firm size and the level of analyst coverage is supported by the empirical findings of Bhushan (1989), Hong, Lim and Stein (2000) and Rock, Sedo and Willenborg (2001). In particular, Hong, Lim, and Stein (2000) examine nine models of cross-sectional determinants of analyst coverage and report that firm size alone yields a R^2 of 0.61, while other variables such as the book-to-market ratio, stock price, industry dummies, stock option dummies, trading turnover and lagged stock returns produce only a marginal increase in R^2.

monitoring other stocks. The attention of traders and arbitrageurs are shifted to stocks that are experiencing turbulent price movements because the opportunity cost of not monitoring those stocks is high. Resultantly, θ rises, and the probability of being picked off falls. The stock's market makers become less sensitive to the effect of monitoring cost on quote updating strategies. Since the monitoring cost function is convex, and with a constant scaling parameter c, a rising θ implies that market makers of large stocks will assign disproportionately *lesser* labor to study the news flows than market makers of small stocks. Thus, news monitoring in large stocks will become less attractive than news monitoring in small stocks, which in turn causes a lesser likelihood of observing a correlation breakdown in large stocks than that in small stocks. To show this mathematically, we differentiate $\partial\sigma^*/\partial c$ with respect to θ:

$$\begin{aligned} \frac{\partial^2\sigma^*}{\partial c \partial\theta} &= -\left[\frac{4}{\theta^3\alpha(1-\alpha)\beta_f}\right]\left[\alpha\theta + 2(1-\alpha)\beta_f\left(2 - \frac{\beta_s}{\beta_f}\theta\right)\right] \\ &< 0. \end{aligned} \tag{20}$$

This gives us the third hypothesis:

***Hypothesis* 3**: *The Size Effect weakens when the opportunity cost of not monitoring other stocks is high.*

Since the opportunity cost of not monitoring other stocks cannot be observed directly, we test the hypothesis using the absolute stock market index return as a proxy. During volatile market circumstances, arbitrageurs, news observers and traders would shift their attention to market-wide news and pay less attention to individual stocks for economic and psychological reasons. Existing evidence shows high return and liquidity correlation across individual stocks in periods of high market volatility (Ang and Chen (2002), Hameed, Kang, and Viswanathan (2010), Karolyi, Lee, and van Dijk (2012)), inferring the relative importance of aggregate market news to individual stock events during market stress. Hypothesis 3 states that the impact of firm-specific news on correlation breakdown is weaker for large stocks when the overall market is experiencing turbulent movement.

3. Data and Variable Construction

In this section, we describe the data sources and the construction of variables used to empirically test our theoretical predictions. Our empirical analysis requires data on the intraday bid and ask quotes of futures and the corresponding underlying stocks, public media news coverage, and stock characteristics.

3.1. Measuring Return Correlations Between Spot and Futures Markets

We collect the U.S. SSF contract data from the Thomson Reuters Tick History (TRTH) database supplied by SIRCA. The TRTH database specializes in providing millisecond time-stamped trades and quotes data for a wide spectrum of stocks and derivatives traded around the world. For each SSF contract, intraday best bid and ask quotes are obtained from the TRTH database. Using this database, we identify the underlying stock and extract the corresponding best bid and ask quotes by matching its stock ticker and the time stamp with those of the SSF contract recorded by TRTH. Due to the availability of the SSF quote data, our sample period spans from 1 July 2009 when SSF quote data became available in the TRTH database through 19 March 2010, which was the expiring date for all March 2010 SSF contracts. We include into our sample all the SSF contracts which expired before or on 19 March 2010 and we do not roll over any SSF contracts.

To construct the conditional correlation measures between spot and futures quote returns, for each day, we gather 10-minute snapshots of the best bid and ask quotes for the SSF and the corresponding spot from 9:30 to 16:00 New York Time. In this study, we employ two different measures of quote return correlation. The first set of conditional correlation measures on the bid and ask quotes, CC_{bid} and CC_{ask} respectively, are based on the unadjusted raw logarithmic returns of the futures and spot quotes. We apply a rolling forward window of six 10-minute intervals. For example, the conditional correlation of bids at 9:40 is computed based on the returns of futures and spot bid quote at 9:40, 9:50, 10:00, 10:10, 10:20 and 10:30.[10] The second set of return correlation measures, CC_{bid}^{BEKK} and CC_{ask}^{BEKK}, are based on the residuals from the vector error correction-BEKK generalized autoregressive conditional heteroskedasticity (VEC-BEKK) model. The description of the VEC-BEKK model is provided in Appendix 1.

For both sets of conditional correlations, we take a daily average of the intraday rolling window correlation values for each stock. These daily average values are used in the subsequent panel regression estimations.

3.2. Measuring Public News Arrivals

Our news coverage data are drawn from the RavenPack database. This database gathers, analyzes and quantifies the real-time business news stories from a wide range of newswires such as Dow Jones Newswires, regional editions of the Wall Street Journal and Baron's. This database uses proprietary algorithms to analyze relevant textual articles from various newswires and produces a range of news measurements including news relevance scores and news sentiment scores. Given its coverage and unique news analysis, this database has been frequently used in the recent literature (e.g., Sabherwal, Sarkar, and Zhang (2011); Kolasinski, Reed, and Ringgenberg (2013); and Shroff, Sun, and White, Zhang (2013)).[11]

[10] To check the robustness of our results, we test alternative time interval lengths such as 15-minute and 30-minute. Our results remain qualitatively unchanged.

[11] Particularly, Shroff, Sun, and White, Zhang (2013) validate the coverage of Ravenpack using a simple test. They examine the overlap in the monthly press release frequency between RavenPack and Factiva for a random sample of 50 firms and find that the press release frequency correlation between the two news databases are as high as 94.7%.

The RavenPack database is essential for this study because it provides both the quantity and quality measure of news, covering the two important dimensions of public news coverage. To measure the quantity dimension of news arrival, RavenPack generates a real-time news relevance score, which is a numeric ranging from zero to 100. A score of zero means the firm is passively mentioned while a score of 100 means the firm plays a key role in the news story. If the firm is referenced in the main title or headline of the news story, it will receive a relevance score of at least 90. In general, scores above 75 are considered significantly relevant. When calculating the relevance score, RavenPack goes beyond analyzing keywords and considers the context of the news story using automated classifiers to detect the roles that firms play in various specific events (such as acquisitions, legal disputes, announcements of corporate actions and executive changes).[12] Our first news arrival measure, $\#News1$, is the total number of relevant news if the firm-specific news relevance score exceeds 75, and zero otherwise. Our second news arrival measure, $\#News2$, includes all news related to the stock. It sums the news relevance scores recorded over a given trading day scaled by 100. Like the first measure, this measure also captures the magnitude of news relevance.

3.3. Measuring News Uncertainty

To measure the news uncertainty on any given day, for each stock we compute the standard deviation of the sentiment scores of stock-specific news arriving over the day. For each piece of news, RavenPack database provides a set of different sentiment scores that examine the content of each news article and infer the potential impact of the news story. In this study, we will use the composite sentiment score (CSS) which measures the news story's overall sentimental tendency and built by combining five sentiment analytics (PEQ, BEE,

[12]For example, a news story about IBM where the firm is referenced in the headline of the news story receives a minimum value of 90. If the headline of the story is "IBM In Software Pact With Raytheon Unit For Navy Program", then the firms IBM and Raytheon will receive a relevance score of 100 since they both play a key role in the story.

BMQ, BCA and BAM) using an intuitive set of rules while ensuring no sentiment disagreement exists amongst the analytics.[13] Appendix 2 outlines the construction procedures of the news sentiment scores considered in this study. CSS was trained on market data using a portfolio of large cap stocks and evaluating intraday fluctuations to determine "strength" or how positive or negative a story is. It is a numeric between zero and 100, with a higher (lower) than 50 value indicating a positive (negative) news item. By using the standard deviation of the CSSs, σ_{CSS}, as a proxy for news uncertainty, we are able to capture the potential differences in opinions reflected through textual news articles arrived on the same day. One attractive feature of this news uncertainty measure is that since we measure the volatility of sentiment scores rather than the sentiment score per se, our results are less sensitive to the proprietary algorithms the RavenPack adopted to build these scores.[14]

In terms of the reliability of the sentiment score, RavenPack did not change the CSS algorithm since its inception, which implies that the way the score is computed is consistent over time. This is very important because we are using the standard deviation of CSSs as proxy for news uncertainty, and if RavenPack changes the algorithm, the proxy will exhibit time variation which may in fact be caused by measurement error. Furthermore, RavenPack version controls their news analytics feeds so that new events and analytics can be added in any future version of the feed without disrupting users' models that are built and backtested using the current version.[15]

---

[13]For robustness, we also examine the impact of different components of CSS that may affect the quote return correlation. See Section 4.1 for further discussions.

[14]This measure is in stark contrast with quote return volatility, which is a commonly used proxy for information uncertainty. The latter measure is inappropriate because volatility is an *outcome* of the trading process and it is unsuitable to proxy for pure public news uncertainty. In other words, quote return volatility is an endogenous outcome, not an exogenous measure of uncertainty.

[15]We also test the reliability of CSS by comparing it with sentiment scores generated from an alternative textual analysis software, DICTION 6.0. DICTION is a dictionary-based program that counts types of words most frequently encountered in contemporary American public discourse and is designed to capture the linguistic style (i.e., verbal tone) of narratives. This software has been extensively used in recent accounting and finance literature (e.g., Demers and Vega (2010) and Rogers, Buskirk, and Zechman (2011)). For comparison, we pick a company, *Johnson & Johnson*, on 19 Feb 2010, on which day the disapproval of their

3.4. Other Variables

In this study, we also control for three stock characteristics, namely, stock price, stock turnover, and firm size. We obtain these data from the CRSP database. Closing stock trading price is used to measure stock price because of the minimum price increments on quote return. Furthermore, highly priced stocks tend to have large dollar bid-ask spreads, and therefore the quote matching between stock spot and futures markets may be weaker. Hence, we expect stock price to be negatively related to return correlation between spot and futures returns. Additionally, we use stock turnover (measured by the number of shares traded scaled by the number of shares outstanding) to proxy for stock trading activity. Stock turnover is loosely related to θ, as it relates to the intensity of arbitrage activity (see the discussion in section 2.3.1). Differentiating σ^* with respect to θ shows that the lower the θ, the greater the arbitrage activity, the higher the cutoff σ^*, suggesting a positive relation between stock turnover and spot-futures return correlation.

Finally, to control for any other unobservable stock characteristics, we also include in our empirical analysis firm size computed as the natural log of end-of-day market capitalization in thousands.

Our final sample contains 1,485 unique stock-futures pairs and 205,603 stock-day observations after we merge the stock and futures quote data with the stock characteristics. Table 1 presents the mean, median, standard deviation, the 10^{th} and the 90^{th} percentiles of the variables of interest. The mean values of the conditional correlation between stock and futures ask (CC_{ask}) and bid prices (CC_{bid}) are 83.23% and 82.94%, respectively, while those

proposed new drug was announced to the public. There are 34 news articles relating to *Johnson & Johnson* in RavenPack on that particular day. DICTION analyzes each of these 34 news articles and generates an *optimism* score which captures the language highlighting positive entailments, similar to the computation of CSS. Overall, the correlation between CSSs from RavenPack and the optimism scores generated by DICTION is reasonable (around 0.40).

of the BEKK conditional correlation measures tend to be slightly lower (Mean CC_{ask}^{BEKK} = 71.33%; Mean CC_{bid}^{BEKK} = 69.18%). The strong average correlation between futures and their underlying stock returns indicates the close market movement across these two markets and potentially algorithmic quote matching between market makers in these two markets.

Another noteworthy observation from Table 1 is the prevalence of public news arrivals among our sample stocks on a single trading day. For an average stock, the mean values of $\#News1$ and $\#News2$ are 1.68 and 2.61, respectively. Moreover, there are some variabilities in within-day news sentiments as the mean value of the standard deviation of news sentiment, σ_{CSS} is 2.53. Finally, we note that the 10th and 90th percentiles of our variables of interest are all within a reasonable range, which rules out the concern of outliers in our subsequent empirical analysis.

<<Insert Table 1 here>>

4. Empirical Analysis

This section discusses the empirical results of the three theoretical predictions in the context of spot-futures return correlation highlighted in Section 2.

4.1. Main Regression results

In the absence of news shock, our model predicts that it is suboptimal for market makers to exert costly efforts in news monitoring. For hedging purposes, they use computerized algorithms to match the futures quotes with the spot quotes. When there is a news shock, and if the size of the shock is sufficiently large, market makers shift from automating the quote matching process to manually monitor and revise the quotes. Hypothesis 1 (News Effect) predicts that there will be a correlation breakdown between spot and futures quote returns. To empirically test this theoretical prediction, we specify the following baseline OLS regression model:

$$y_{i,t} = \alpha + \beta_1 x_{i,t} + \beta_2 Price_{i,t} + \beta_3 Turn_{i,t} + \beta_4 Size_{i,t} + \varepsilon_{i,t}, \tag{21}$$

where $y_{i,t}$ are the unadjusted and adjusted correlations between spot and futures ask and bid quotes: CC_{ask}, CC_{bid}, CC_{ask}^{BEKK} and CC_{bid}^{BEKK} for stock i on day t. $x_{i,t}$ represents two alternative measures of news arrivals: $\#News1$ and $\#News2$. After controlling for firm-level characteristics including stock price, stock turnover and firm size, we expect β_1 to be negative as a greater intensity of news arrival leads to a greater correlation breakdown (i.e. lower CC). Additionally, we include day dummies to address any time series variation.[16] In our model, robust standard errors adjusted for firm-level and time-period clustering are used to account for heteroskedasticity and correlation across firms and time periods (Petersen (2009) and Thompson (2011)).[17]

Table 2 shows a consistent and robust negative relation between two news arrival measures and correlation proxies, indicating that when news arrives, the correlation between spot and futures quote returns breaks down. In Panel A of Table 2, when $\#News1$ is used as the news arrival measure, the coefficients of $\#News1$ are -0.183 (t-value = -5.20) and -0.222 (t-value = -5.79), respectively for CC_{ask} and CC_{bid} as shown in Models 1-2. From an economic perspective, this result suggests that one standard deviation increase in the arrival of relevant news articles leads to 94 (114) basis points reduction in the correlation between spot and futures ask (bid) quote returns.

Using the VEC-BEKK conditional correlations, we document qualitatively similar re-

[16] As for day dummies, we assign a dummy variable for each trading day invovled in our sample period.

[17] According to Thompson (2011), double clustering is important in datasets when the panel regression errors include significant time and firm components. Monte Carlo simulation experiments also show that the simultaneous clustering by firms and time lead to significantly more accurate inference in finance panels. We have also used firm-level clustering only and the statistical significance of our results are similar to the case of double-clustered standard errors.

sults (reported in Models 3 and 4). The coefficients of $\#News1$ continue to be negative and statistically significant at conventional levels. While one would argue that our results could be driven by events occurring on specific days, including day dummies in our baseline regression specification (reported in Models 5-8) does not change our conclusion. Our control variables yield findings that are consistent with our expectations. The spot-futures quote return correlation tends to be higher for low-priced stocks, large stocks, and stocks that are actively traded.

Recall that $\#News2$ aggregates all news relevance scores over trading day t. Therefore, it improves $\#News1$ by including all relevant news (relevance score > 0) instead of restricting news with relevance score exceeding 75. For ease of interpretation, we scale the $\#News2$ measure by 100. We report the results in Panel B of Table 2. Evidently, our conclusion remains unchanged, as coefficients of $\#News2$ are negative and statistically significant at 1% level for all models.

<<Insert Table 2 here>>

It should be noted that previous news measures ($\#News1$ and $\#News2$) only provide the quantity dimension of public news. They do not capture the level of ex ante uncertainty stemming from the public news.[18] To test Hypothesis 1 more directly, we use the standard deviation of CSS of each news story available over the day, σ_{CSS}, as an alternative news proxy. If σ_{CSS} is high, there may be a high degree of heterogeneity or dispersion of opinion among news released over that day. The downside of this measure is that it requires at least two available CSSs for each stock on each trading day to compute a valid standard deviation, which substantially reduces our sample size.

[18]Here ex ante news uncertainty refers to the uncertainty about public news prior to financial market reaction.

Using this new news measure, Table 3 reports consistent evidence in support of Hypothesis 1. The coefficient of σ_{CSS} is negative and statistically significant (Models 1-8) across all model specifications. Notice that the sign and the significance of σ_{CSS} coefficients are not altered even when the magnitude of relevant news measure $\#News2$ is included. These results suggest that both the quantity and quality of news arrivals matter to market participants in the quotes setting process in related asset markets.

<<Insert Table 3 here>>

To gain better insights into the components of news sentiment that may matter to the quote return correlation, we further investigate the effect of news uncertainty using the standard deviation of three specific news sentiment classifiers. These specific news sentiment classifiers are BCA, PEQ and BMQ.[19] BCA sentiment score focuses on news reports on corporate action announcements; PEQ tracks positive and negative words and phrases in articles about global equities; and BMQ specializes in analyzing the sentiment of short commentary and editorials on global equities. After controlling for the number of public news ($\#News2$) as well as other firm-specific variables, we continue to find negative and statistically significant coefficients on these three specific news sentiment scores (i.e., σ_{BCA}, σ_{PEQ}, and σ_{BMQ}) across all model specifications in Table 4.

<<Insert Table 4 here>>

In summary, we document a statistically significant relationship between the amount and volatility of news arrivals and return correlation breakdown across stock spot and futures

[19] It should be noted that the RavenPack database provides more than these three news sentiment classifiers. For example, it offers information about sentiments classifiers on news impact projection, earnings evaluation and releases. To best suit our research purpose, we choose general classifiers that are applicable to all stocks and refrain from using those that are either forward-looking or narrowly cover limited aspects of corporate actions.

markets. This finding indicates that the arrival of public news that yields an immediate disagreement among traders will prompt the market makers to switch from automatic quote updating to monitoring the news flow, which eventually results in the return correlation breakdown in two seemingly closely related markets.

4.2. Cost of Monitoring, Public News Arrivals, and Correlation Breakdown

Hypothesis 2 (Size Effect) addresses the role of firm size in the relationship between public news arrivals and correlation breakdown. Our model predicts that large firms require lower monitoring efforts than smaller stocks in understanding the news shocks, and this leads to more negative news effect on the spot-futures return correlation for these stocks. We test this hypothesis by augmenting the baseline equation specification to include a term that interacts firm size with the public news arrival proxy. If our theoretical prediction is correct, we should expect the interaction to be negative.

The empirical results outlined in Table 5 are in line with our expectation. Take Model 1 for example. Without controlling for the interaction term $\#News2 \times Size$, the coefficient of $\#News2$ is -0.142 as shown in Model 1, Panel B of Table 2. While the $\#News2$ coefficient becomes positive after we include the interaction term in Model 1 of this table, the coefficient of $\#News2 \times Size$ is -0.071 and it is statistically significant at the 1% level. We obtain a similar set of results across different measures of conditional correlation measures in Models 2-4. When the VEC-BEKK model is used to account for the long-run dynamic relationship between the two series, the magnitudes of the coefficients on both $\#News2$ and its interaction term increase by at least twofold, along with jump in R^2 from around 0.27 to 0.44. Models 5-8 report results that replace $Size$ with a dummy variable, $Large$, which equals one if the stock belongs to the top size quartile and zero otherwise. The conclusion remains unchanged as the sum of the coefficients of $\#News2$ and $\#News2\times Large$ are consistently negative.

<<Insert Table 5 here>>

Overall, our results support the notion that the observed negative relation between public news arrivals and spot-futures return correlation are mainly driven by large firms, which could be due to the fact that large firms are more informationally transparent and thus incur less manual monitoring costs.

4.3. Opportunity Cost of Not Monitoring Other Stocks, Public News Arrivals, and Correlation Breakdown

Hypothesis 3 implies that the firm size effect on the relation between spot-futures return correlation and public news arrival weakens when the overall market is experiencing turbulent movement. Prior studies suggest that when the market is undergoing extreme movements (especially during market downturns), individual investors tend to be less concerned about individual firms' performance but base their investment decisions on the whole market performance. In the context of our model, during turbulent market condition, market makers are less sensitive to the impact of monitoring cost to quote updating strategies because they face a lower risk of being picked off, and hence the effect of firm size on news-correlation is weaker.

To test Hypothesis 3, we consider a three-way interaction term. We interact the absolute return of the S&P500 Index, $MVol$, with the interaction term: $\#News2\times Size$. Results are reported in Table 6. In all models, the conditional correlation falls when the overall market experiences an extreme movement, as the coefficient of $MVol$ is negatively and statistically significant at the 1% level. The coefficient of $\#News2\times MVol$ in Models 1, 3, 5 and 7 (without the three-way interaction) is positive and significant, indicating that firm-specific news arrivals matter more for cross-market traders when overall stock market is less turbulent, though its significance falls when CC_{ask}^{BEKK} and CC_{bid}^{BEKK} are used.

After including the three-way interaction term, $\#News2{\times}Size{\times}MVol$, in Models 2, 4, 6 and 8, we observe that the coefficients on the three-way interaction term are positive in all models. Moreover, this coefficient is only statistically insignificant in Model 8. These findings support our conjecture that, when faced with extreme market-wide movements, market makers become less sensitive to the impact of monitoring cost on quote updating strategies as the attention of newswatchers and arbitrageurs is shifted to market-wide news.

<<Insert Table 6 here>>

4.4. Robustness Check

While we observe a significant relationship between public news arrival and the return correlation breakdown between spot and the underlying futures, this result might be driven by the selection of the news arrival proxy and/or the use of regression estimation method. We specifically address these concerns in this section.

Da, Engelberg, and Gao (2011) propose a new and direct measure of investor attention using the search frequency in Google (i.e., Search Volume Index (SVI)). We do not use this measure as our primary proxy for public news arrival for two reasons. One, this measure is likely to capture the revealed public attention from retail investors, which is endogenously determined. Retail investors may search for news because the textual content of the news itself or the news headlines have sentimental tendency. Second, this measure ignores one important dimension of news effect, that is, the differential opinions among investors. Nonetheless, to confirm the robustness of our results, we gather the Google search frequency data from Google Trends on a daily basis using stock tickers and denote this variable as SVI.

We re-estimate our main regression models by substituting the public news arrival proxy, $\#News2$ with SVI. As shown in Models 1-4 of Table 7 Panel A, we continue to find a negative association between the conditional correlation measures and SVI, which is consistent with our prior findings. In Models 5-8, we further observe a negative and statistically significant coefficient of $SVI \times Size$ when the cross-asset correlation is measured by adjusted conditional correlations. This is in line with the idea that the news effect on the correlation breakdown is most pronounced for large firms even when using an alternative news arrival measure. In Models 9-12, the positive coefficients on $SVI \times Size \times MVol$ support the notion that the spot-futures return correlations of large firms are less affected by firm-specific news when the entire stock market is exposed to volatile trading conditions. Taken together, the results using SVI as the alternative proxy for news arrival are consistent with our previous findings.

Since all of our dependent variables are bounded between -100 and 100, one might argue that OLS regression estimators are inconsistent. We re-estimate the main regression model using the Tobit regression model. Results are reported in Panel B of Table 7. Our overall conclusion remains unchanged as results are qualitatively similar to those estimated by the OLS model.

<<Insert Table 7 here>>

5. Conclusion

In this study, we focus on a specific form of algorithmic trading strategy that enables traders to automate the process of matching quotes across asset markets. We postulate that when public news arrives, the trade-offs between the costs and the benefits of algorithmic trading and human monitoring have important implications on the cross-asset return correlation. In particular, we develop a simple model that shows how the return correlation between assets breaks down when public news arrives, especially when the news impact is highly uncertain ex ante.

Our model predicts that futures market makers may switch from algorithmic quote matching to monitoring and studying the news feed directly when the news content is vague about the asset value. While the latter requires costly labor processing and human attention, futures market makers compensate the cost by setting a wider spread, which then causes a momentary breakdown in spot-future return relation. Our model suggests that the likelihood of the breakdown varies across assets depending on firm size and market conditions, and they relate respectively to the cost of news monitoring and the opportunity cost of not monitoring other stocks.

To test our theoretical predictions, we study the quote return correlation between the stock and its SSF contract. Our main findings show that the spot-futures return correlation decreases with both the number of public news arrivals and the ex ante news uncertainty, proxied by the volatility of the sentiment scores related to the news released. These results are consistent with the notion that the uncertainty faced by traders in two closely related markets rises with the amount of public news arrivals, and the increase in the amount of public information forces the traders to switch from algorithmic quote matching to manual monitoring.

We further investigate how certain stock characteristics affect the association between public news arrivals and cross-market return correlation. Our empirical evidence shows that the negative effect of public news arrivals on spot-futures return correlation tends to be more pronounced for larger stocks compared with smaller stocks. Using the volatile stock market condition as a proxy for opportunity cost of not monitoring other stocks, we show that the effect of firm size on the news-correlation sensitivity falls when the overall stock market is going through a turbulent period.

This study cautions us against the vulnerability of algorithmic trading upon public news arrivals. Importantly, our findings pinpoint the reasoning for and empirical existence of human interference in the algorithmic trading process when public news arrives and when the news is too complex for machine interpretation. Even though the scope of this study is set on one algorithmic trading strategy, the implications of this study can be extended to generic algorithmic trading activities in future research.

Appendix 1: VEC-BEKK Model

The details of the VEC-BEKK Model are explained in this appendix. This model is used because it captures the cointegrating properties of futures and spot prices due to their cost of carry relation. Define Δf_t and Δs_t as the log return of futures and spot quotes, respectively. The VEC model is written as follows:

$$\Delta s_t = \gamma_s z_{t-1} + \sum_{j=1}^{3} \phi_{ss,j} \Delta s_{t-j} + \sum_{j=1}^{3} \phi_{sf,j} \Delta f_{t-j} + \varepsilon_{s,t-1}$$

$$\Delta f_t = \gamma_f z_{t-1} + \sum_{j=1}^{3} \phi_{fs,j} \Delta s_{t-j} + \sum_{j=1}^{3} \phi_{ff,j} \Delta f_{t-j} + \varepsilon_{f,t-1}$$

where z_{t-1} is the error correction term at time $t-1$, for $z_{t-1} = f_t - \alpha_0 - \alpha_1 s_t - \rho t$. We include the time trend, t, in the cointegrating equation to allow for the time varying cost of carry term.[20] Our Johansen (rank) test results suggest that the futures and spot quotes series are cointegrated. Based on the Akaike Information and Schwarz Bayesian Information Criteria (AIC and SBIC), we choose three lags for our VEC model specification.

To compute the conditional correlations, we adopt the diagonal version of the BEKK GARCH model proposed by Engle and Kroner (1995). According to Chang, McAleer, and Tansuchat (2011), the diagonal BEKK is the best model for minimum-variance hedging:

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{s,t} \\ \varepsilon_{f,t} \end{pmatrix} \Bigg| Q_{t-1} \sim Student-t(0, H_t, v), \ H_t \equiv \begin{pmatrix} h_{s,t} & h_{sf,t} \\ h_{sf,t} & h_{f,t} \end{pmatrix}$$

where:

$$h_{s,t} = M_1 + A_1^2 \varepsilon_{s,t-1}^2 + B_1^2 h_{s,t-1}^2,$$

$$h_{f,t} = M_1 + A_2^2 \varepsilon_{f,t-1}^2 + B_2^2 h_{f,t-1}^2,$$

$$h_{sf,t} = M_3 + A_1 A_2 \varepsilon_{s,t-1} \varepsilon_{f,t-1} + B_1 B_2 h_{sf,t-1}^2.$$

[20] According to the cost of carry relation, $F_t = S_t e^{c(T-t)}$, where c is the cost of carry. F_t and S_t are the futures and spot quotes at time t. Taking log on both sides yields: $ln(F_t) = ln(S_t) + c(T-t)$.

The model assumes that the error terms $\varepsilon_t = \begin{pmatrix} \varepsilon_{s,t} \\ \varepsilon_{f,t} \end{pmatrix}$ in the VEC equations follow a multivariate conditional Student-t distribution, which is preferred because it can capture the leptokurtic property of spot and futures returns. Q_{t-1} is the information set up to time $t-1$, H_t is the conditional-variance matrix of the diagonal BEKK model at time t, and v is the degrees of freedom of the Student-t distribution. The conditional variance of the spot ($h_{s,t}$) and futures ($h_{f,t}$) returns and their conditional covariance ($h_{sf,t}$) are specified as a GARCH(1,1) structure. To calculate the conditional correlations between the spot and futures returns, we apply the following equation:

$$Corr_{sf,t} = \frac{h_{sf,t}}{\sqrt{h_{s,t}h_{f,t}}}$$

The conditional variance-covariance matrix H_t is estimated by maximizing the following log-likelihood function $L(\Theta)$ with respect to the vector of the estimated parameters (Θ) in H_t:

$$L(\Theta) = \sum_{t=1}^{T} ln\{l_t(\Theta)\},$$

where:

$$l_t(\Theta) = \frac{\Gamma[(2+v)/2]}{\Gamma(v/2)[\pi(v-2)]}|H_t|^{-1/2}\left(1 + \frac{1}{v-2}(\varepsilon_t^T H_t^{-1}\varepsilon_t)\right)^{-(2+v)/2}$$

$\Gamma(\bullet)$ is the gamma function and T is the total number of observations in the spot and futures returns series.

Appendix 2: Methodologies of RavenPack News Analytics - Dow Jones Edition

This appendix outlines the main steps adopted by the RavenPack database for calculating the various news sentiment scores (CSS, BCA, PEQ and BMQ) used in this study.

Market Response Methodology

RavenPack's Market Response methodology underpins the CSS and is based on a Rule Base that identifies and maps individual words or word combinations in the story headline to the price impact on stocks of companies mentioned in the headline. The price impact is measured in the hours ahead of the arrival of the news item and is transformed into an impact score using advanced machine learning techniques.

Step 1: A Classification Base is defined in the first step. This defines the types of stories that contain the content relevant for tagging.

Step 2: In the second step, a large sample is analyzed to create a Rule Base. A sample set of stories in the Classification Base developed in step 1 is drawn from RavenPack's news database for a fixed date range. The headlines of these stories are extracted and parsed into words to form a list of candidates of individual words and word combinations that are typical for such headline stories.

Step 3: The third step creates an impact score using the Rule Base. Different advanced machine learning techniques are applied with the objective of creating an impact score that identifies the probability of the volatility of a particular stock to be either higher or lower than the volatility of the market.

Step 4: Historical analysis is generated and real-time tagging is enabled. This process involves several consistency checks of historical data and generation of volume statistics.

When this process is complete, the series is published.

Step 5: Classifiers are re-evaluated on a quarterly basis. This process involves completing step 2 for stories sampled outside of the date range of the original sample or most recent quarterly re-evaluation. If the accuracy level is 10 percent lower than the level when the series was originally released, a new series is developed.

Traditional Tagging Methodology

RavenPack's traditional tagging methodology underpins the PEQ score and is based on a Rule Base that maps specified words, phrases, combinations and other word-level definitions to pre-defined sentiment values. Each story may match several of these rules. Two Rule Bases are defined: one that identifies positive sentiment and another that identifies negative sentiment.

Step 1: A Classification Base is defined in the first step. This defines the types of stories that contain the content relevant for tagging. The ideal Classification Base contains only stories that contain news that affect the target market or asset class.

Step 2: A large sample is analyzed to create a Rule Base. Experts read and extract key language from a randomly selected sample. This can include individual words, phrases, template phrases and complex linguistic rules that account for diverse types of language.

Step 3: The Rule Base is tested on a large sample. A Rule Base undergoes extensive review and expansion through testing on large sample sets. During this process, methodological consistency is maintained by not applying any weights or sentiment values to individual stories.

Step 4: This step involves generating historical analysis and enabling real-time tagging. The process includes several consistency checks of historical data and generation of volume statistics. When this process is complete, the series is published.

Step 5: In this step, classifiers are re-evaluated on a quarterly basis. This process involves completing step 3 for stories sampled outside of the date range of the original sample or most recent quarterly re-evaluation. Experts who analyze the original Rule Base carefully search for new language patterns that are not detected. If the original Rule Base expands by more than 10 percent as a result of this process, a new series is developed.

Expert Consensus Tagging Methodology

RavenPack's expert consensus methodology underpins the BCA and BMQ scores and entails a group of financial experts manually tagging a set of stories that is later used as a basis for automated computer classification using a Bayes classifier.

Step 1: A Classification Base is defined in the first step. This defines the types of stories that contain the content relevant for tagging.

Step 2: Experts build an internal tagging guide in the second step. A team of in-house experts with extensive backgrounds in linguistics, finance and economics first develop and agree upon a set of parameters and basic assumptions that will guide sentiment tagging.

Step 3: In the third step, a large sample is tagged. A sample set of stories in the Classification Base developed in step 1 is drawn from RavenPack's news database for a fixed date range. Stories are randomly selected for tagging. A group of experts read and classify the sample using the tagging guide developed in step 2.

Step 4: The software is trained from sample to automate tagging. A Bayes classifier uses supervised learning to discern patterns in expert tagging and establish rules for future automation. This automated tagging process must meet exceptional levels of accuracy in order to be made available to clients. In cases when accuracy is not sufficiently high, step 3 is repeated with a larger sample set.

Step 5: Historical analysis is generated and real-time tagging is enabled. This process involves several consistency checks of historical data and generation of volume statistics. When this process is complete, the series is published.

Step 6: Classifiers are re-evaluated on a quarterly basis. This process involves completing step 3 for stories sampled outside of the date range of the original sample or most recent quarterly re-evaluation. The results of this expert classification are compared to the results of automated classification. If the accuracy level is 10 percent lower than the level when the series was originally released, a new series is developed.

References

Ang, A., and J. Chen. "Asymmetric Correlations of Equity Portfolios." *Journal of Financial Economics*, 63 (2002), 443-494.

Antweiler, W., and M. Z. Frank. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance*, 59 (2004), 1259-1294.

Bhushan, R. "Firm Characteristics and Analyst Following." *Journal of Accounting and Economics*, 11 (1989), 225-274.

Biais, B.; T. Foucault; and S. Moinas. "Equilibrium High Frequency Trading." Working Paper, Toulouse School of Economics and HEC Paris (2011).

Cao, H., and H. Ou-Yang. "Differences of Opinion of Public Information and Speculative Trading in Stocks and Options." *Review of Financial Studies*, 22 (2009), 299-335.

Chaboud, A.; B. Chiquoine; E. Hjalmarsson; and C. Vega. "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market." International Finance Discussion Papers 980. Board of Governors of the Federal Reserve System (U.S.) (2013).

Chakrabarty, B., and P. Moulton. "Earnings Announcements and Attention Constraints: The Role of Market Design." *Journal of Accounting and Economics* 53 (2012), 612-634.

Chang, C.; M. McAleer; and R. Tansuchat. "Crude Oil Hedging Strategies using Dynamic Multivariate GARCH." *Energy Economics* 33 (2011), 912-923.

Cvitanic, J., and A. Kirilenko. "High Frequency Traders and Asset Prices." SSRN Working Paper (2010).

Da, Z; J. Engelberg; and P. Gao. "In Search of Attention." *Journal of Finance*, 66 (2011), 1461-1499.

Dechow, P.M., and I.D. Dichev. "The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors." *The Accounting Review* (Supplement), 77 (2002), 35-39.

Demers, E., and C. Vega. "Soft Information in Earnings Announcements: News or Noise?" Working Paper, University of Virginia and Federal Reserve System (2010).

Donefer, B. "Algos Gone Wild: Risk in the World of Automated Trading Strategies." *Journal of Trading* 5 (2010), 31-34.

Engle, R., and K. Kroner. "Multivariate Simultaneous Generalized ARCH." *Econometric Theory* 11 (1995), 122-150.

Foucault, T.; O. Kadan; and E. Kandel. "Liquidity Cycles and Make/Take Fees in Electronic Markets." *Journal of Finance*, 68 (2013), 299-341.

Foucault, T.; A. Röell; and P. Sandås. "Market Making with Costly Monitoring: An Analysis of the SOES Controversy." *Review of Financial Studies*, 16 (2003), 345-384.

Gadarowski, C. "Financial Press Coverage and Expected Stock Returns." Working Paper, Cornell University (2002).

Gerig, A., and D. Michayluk. "Automated Liquidity Provision and the Demise of Traditional Market Making." Working Paper, University of Technology, Sydney (2010).

Gomber, P.; B. Arndt; M. Lutat; and T. Uhle. "High-Frequency Trading." Working Paper, University of Frankfurt and Goethe University of Frankfurt (2011).

Hameed, A.; W. Kang; and S. Viswanathan. "Stock Market Declines and Liquidity." *Journal of Finance*, 65 (2010), 257-293.

Harris, M., and A. Raviv. "Differences of Opinion Make a Horse Race." *Review of Financial Studies*, 6 (1993), 473-506.

Hasbrouck, J., and G. Saar. "Low-Latency Trading." *Journal of Financial Markets* 16 (2013), 646-679.

Hendershott, T.; C. Jones; and A. Menkveld. "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance*, 66 (2011), 1-33.

Hong, H.; T. Lim; and J. Stein. "Bad News Travel Slowly: Size, Analyst Coverage and the Profitability of Momentum Strategies." *Journal of Finance*, 55 (2000), 265-295.

Kolasinski, A. C.; A. V. Reed, and M. C. Ringgenberg. "A Multiple Lender Approach to Understanding Supply and Search in the Equity Lending Market." *Journal of Finance*, 68 (2013), 559-595.

Jarrow, R.; and P. Protter. "A Dysfunctional Role of High Frequency Trading in Electronic Markets." *International Journal of Theoretical and Applied Finance*, 15 (2012), 1-15.

Jovanovic, B., and A. Menkveld. "Middlemen in Limit-Order Markets." Yale University Working Paper (2012).

Kahneman, D. *Attention and Effort*. New Jersey: Prentice Hall (1973).

Kandel, E., and N.D. Pearson. "Differential Interpretation of Public Signals and Trade in Speculative Markets." *Journal of Political Economy*, 103 (1995), 831-872.

Karolyi, A.; K.-H. Lee; and M.A. van Dijk. "Understanding Commonality in Liquidity around the World." *Journal of Financial Economics* 105 (2012), 82-112.

Liu, W. M. "Monitoring and Limit Order Submission Risks." *Journal of Financial Markets*, 12 (2009), 107-141.

Petersen, M. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies* 22 (2009), 435-480.

Riordan, R.; A. Storkenmaier; M. Wagener; and S. Zhang. "Public Information Arrival: Price Discovery and Liquidity in Electronic Limit Order Markets." *Journal of Banking and Finance*, 37 (2013), 1148-1159.

Rock, S.; S. Sedo; and M. Willenborg. "Analyst Following and Count-Data Econometrics." *Journal of Accounting and Economics*, 30 (2001), 351-373.

Rogers, J.; A. Buskirk; and S. Zechman. "Disclosure Tone and Shareholder Litigation." *The Accounting Review*, 86 (2011), 2155-2183.

Sabherwal, S.; S. K. Sarkar; and Y. Zhang. "Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News." *Journal of Business Finance and Accounting*, 38 (2011), 1209-1237.

Shroff, N.; A. X. Sun; H. D. White; and W. Zhang. "Voluntary Disclosure and Information Asymmetry: Evidence from the 2005 Securities Offering Reform." *Journal of Accounting Research*, 51 (2013), 1299-1345.

Tetlock, P. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance*, 62 (2007), 1139-1168.

Thompson, S. "Simple Formulas for Standard Errors that Cluster by Both Firm and Time." *Journal of Financial Economics* 99 (2011), 1-10.

Table 1
Summary Statistics

This table presents the summary statistics of all the variables employed in this study. The summary statistics include mean value (Mean), median value (Median), standard deviation (SD), 10th percentiles (P10) and 90th percentiles (P90) for each variable. CC_{ask} (CC_{bid}) is the conditional correlation (in %) between spot and futures ask (bid) quote log returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC_{ask}^{BEKK} (CC_{bid}^{BEKK}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. $\#News1$ is the number of times the news relevance score exceeds 75 on a given trading day (this variable is assumed zero if the relevance score falls below 75). $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. σ_{CSS} is the standard deviation of composite sentiment scores (CSS) available over the day. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. $MVol$ is the absolute percentage return of the S&P500 Index. The number of stock-day observations (Nobs) varies depending on data availability. The sample period is from 20 July 2009 to 19 March 2010.

Variable	Nobs	Mean	Median	SD	P10	P90
Correlation Coefficients						
CC_{ask}	205,603	83.23	94.71	22.36	48.62	99.26
CC_{bid}	205,603	82.94	94.47	22.72	48.29	99.23
CC_{ask}^{BEKK}	201,683	71.33	94.58	35.35	13.90	99.39
CC_{bid}^{BEKK}	201,683	69.18	91.59	36.02	11.43	99.25
News Variables						
$\#News1$	205,603	1.68	0.00	5.13	0.00	4.00
$\#News2$	205,603	2.61	0.41	8.17	0.00	5.64
σ_{CSS}	94,829	2.53	1.90	2.76	0.00	5.88
Control Variables						
$Price$	205,603	28.79	22.10	31.31	5.35	56.75
$Turn$	205,603	0.02	0.01	0.02	0.00	0.03
$Size$	205,603	14.65	14.68	1.74	12.43	16.90
$MVol$	205,603	0.75	0.63	0.56	0.10	1.60

Table 2
Public News Arrivals and Correlation Breakdown

This table presents the regression results of several conditional correlation estimates between spot and the underlying futures on news arrival proxies and other control variables. Conditional correlation proxies, alternatively, are CC_{ask}, CC_{bid}, CC_{ask}^{BEKK}, and CC_{bid}^{BEKK}. CC_{ask} (CC_{bid}) is the conditional correlation between spot and futures ask (bid) log returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC_{ask}^{BEKK} (CC_{bid}^{BEKK}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. Public news arrivals are proxied by either $\#News1$ or $\#News2$. $\#News1$ is the number of times the news relevance score exceeds 75 on a given trading day (this variable is assumed zero if all the relevance scores fall below 75 on a given trading day). $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. t-statistics shown in the parentheses are based on robust standard errors adjusted for both time clustering (by day) and firm-level clustering. ***, **, and * indicate 1%, 5% and 10% statistical significance, respectively. All the reported coefficients are scaled upward by 100. Nobs is the number of stock-day observations. The sample period is from 20 July 2009 to 19 March 2010.

Panel A: Public News Arrival Proxy: $\#News1$

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#News1$	-0.183***	-0.222***	-0.314***	-0.379***	-0.158***	-0.191***	-0.300***	-0.369***
	(-5.20)	(-5.79)	(-4.82)	(-4.86)	(-4.62)	(-5.12)	(-4.67)	(-4.75)
$Price$	-0.040***	-0.034***	-0.013	0.012	-0.040***	-0.035***	-0.011	0.013
	(-3.61)	(-3.24)	(-0.85)	(0.58)	(-3.58)	(-3.26)	(-0.74)	(0.66)
$Turn$	86.736***	80.387***	184.267***	175.021***	88.305***	81.233***	183.202***	172.751***
	(5.71)	(5.44)	(6.40)	(6.39)	(5.62)	(5.32)	(6.26)	(6.20)
$Size$	7.062***	7.108***	13.862***	14.299***	6.943***	7.007***	13.719***	14.176***
	(23.07)	(23.02)	(38.09)	(38.66)	(23.42)	(23.47)	(38.23)	(38.59)
Constant	-20.090***	-21.055***	-134.244***	-143.284***	-14.811***	-16.054***	-125.076***	-136.409***
	(-4.10)	(-4.26)	(-24.10)	(-26.12)	(-3.40)	(-3.65)	(-23.81)	(-26.05)
Day dummies	No	No	No	No	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.270	0.266	0.437	0.459	0.361	0.353	0.460	0.478
Nobs	205,603	205,603	201,683	201,683	205,603	205,603	201,683	201,683

Panel B: Public News Arrival Proxy: $\#News2$

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#News2$	-0.142***	-0.173***	-0.249***	-0.301***	-0.128***	-0.157***	-0.242***	-0.296***
	(-4.56)	(-5.11)	(-4.72)	(-4.52)	(-4.25)	(-4.85)	(-4.63)	(-4.48)
$Price$	-0.040***	-0.034***	-0.013	0.012	-0.039***	-0.035***	-0.011	0.014
	(-3.59)	(-3.22)	(-0.85)	(0.63)	(-3.56)	(-3.24)	(-0.73)	(0.70)
$Turn$	86.579***	80.240***	184.232***	174.962***	88.441***	81.494***	183.410***	172.951***
	(5.71)	(5.43)	(6.42)	(6.39)	(5.63)	(5.32)	(6.27)	(6.21)
$Size$	7.108***	7.164***	13.951***	14.406***	6.989***	7.066***	13.809***	14.284***
	(23.11)	(23.06)	(38.40)	(39.27)	(23.48)	(23.52)	(38.58)	(39.26)
Constant	-20.702***	-21.812***	-135.438***	-144.719***	-15.420***	-16.827***	-126.259***	-137.840***
	(-4.21)	(-4.39)	(-24.34)	(-26.55)	(-3.53)	(-3.80)	(-24.09)	(-26.55)
Day dummies	No	No	No	No	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.271	0.267	0.439	0.460	0.362	0.354	0.462	0.480
Nobs	205,603	205,603	201,683	201,683	205,603	205,603	201,683	201,683

Table 3
Public News Uncertainty and Correlation Breakdown

This table presents the regression results of several conditional correlation estimates between spot and the underlying futures on news arrival proxies and other control variables including day dummies. Conditional Correlation Proxies, alternatively, are CC_{ask}, CC_{bid}, CC_{ask}^{BEKK}, and CC_{bid}^{BEKK}. CC_{ask} (CC_{bid}) is the conditional correlation between spot and futures ask (bid) log returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC_{ask}^{BEKK} (CC_{bid}^{BEKK}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. σ_{CSS} is equal to the standard deviation of composite sentiment scores available over the day. $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. t-statistics shown in the parentheses are based on robust standard errors adjusted for both time clustering (by day) and firm-level clustering. ***, **, and * indicate 1%, 5% and 10% statistical significance, respectively. All the reported coefficients are scaled upward by 100. Nobs is the number of stock-day observations. The sample period is from 20 July 2009 to 19 March 2010.

	CC_{ask}	CC_{ask}	CC_{bid}	CC_{bid}	CC_{ask}^{BEKK}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
σ_{CSS}	-0.384***	-0.295***	-0.423***	-0.312***	-0.541***	-0.383***	-0.649***	-0.463***
	(-5.75)	(-4.63)	(-6.51)	(-5.12)	(-5.28)	(-3.90)	(-5.79)	(-4.55)
$\#News2$		-0.094***		-0.118***		-0.168***		-0.198***
		(-3.63)		(-4.15)		(-4.08)		(-3.60)
$Price$	-0.034***	-0.034***	-0.027**	-0.027**	-0.024*	-0.025**	-0.001	-0.001
	(-2.92)	(-2.88)	(-2.46)	(-2.45)	(-1.93)	(-2.09)	(-0.07)	(-0.09)
$Turn$	47.239***	50.850***	40.949***	45.473***	90.353***	97.145***	82.895***	90.907***
	(3.97)	(4.19)	(3.44)	(3.76)	(4.57)	(4.82)	(4.29)	(4.68)
$Size$	5.869***	6.002***	5.772***	5.940***	10.711***	10.955***	11.015***	11.303***
	(16.51)	(16.95)	(15.89)	(16.46)	(23.09)	(24.01)	(22.44)	(24.19)
Constant	-0.663	-2.504	0.366	-1.940	-82.111***	-85.491***	-89.386***	-93.373***
	(-0.12)	(-0.45)	(0.06)	(-0.34)	(-11.14)	(-11.77)	(-11.64)	(-12.70)
$\bar{R}^2$	0.241	0.244	0.228	0.232	0.340	0.344	0.349	0.354
Nobs	94,829	94,829	94,829	94,829	93,874	93,874	93,874	93,874

Table 4
Alternative Public News Uncertainty Proxies and Correlation Breakdown

This table presents the regression results of several conditional correlation estimates between spot and the underlying futures on news arrival proxies, and other control variables including day dummies. Conditional Correlation Proxies, alternatively, are CC_{ask}, CC_{bid}, CC_{ask}^{BEKK}, and CC_{bid}^{BEKK}. CC_{ask} (CC_{bid}) is the conditional correlation between spot and futures ask (bid) quote returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC_{ask}^{BEKK} (CC_{bid}^{BEKK}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. σ_{BCA}, σ_{PEQ}, and σ_{BMQ} are, respectively, the standard deviation of BCA, PEQ, and BMQ news sentiment scores over a given trading day. $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. t-statistics shown in the parentheses are based on robust standard errors adjusted for both time clustering (by day) and firm-level clustering. ***, **, and * indicate 1%, 5% and 10% statistical significance, respectively. All the reported coefficients are scaled upward by 100. Nobs is the number of stock-day observations. The sample period is from 20 July 2009 to 19 March 2010.

	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
σ_{BCA}	-0.081***	-0.086***	-0.100***	-0.121***								
	(-4.64)	(-5.02)	(-3.82)	(-4.47)								
σ_{PEQ}					-0.065***	-0.072***	-0.091***	-0.106***				
					(-5.16)	(-5.66)	(-4.18)	(-4.92)				
σ_{BMQ}									-0.032***	-0.035***	-0.044***	-0.050***
									(-3.62)	(-4.12)	(-2.91)	(-3.38)
$\#News2$	-0.088***	-0.111***	-0.162***	-0.190***	-0.098***	-0.121***	-0.171***	-0.203***	-0.106***	-0.130***	-0.183***	-0.217***
	(-3.40)	(-3.92)	(-3.92)	(-3.45)	(-3.75)	(-4.24)	(-4.16)	(-3.68)	(-4.01)	(-4.51)	(-4.40)	(-3.89)
$Price$	-0.034***	-0.026**	-0.024**	-0.001	-0.034***	-0.027**	-0.025**	-0.001	-0.035***	-0.027**	-0.025**	-0.002
	(-2.89)	(-2.45)	(-2.07)	(-0.05)	(-2.86)	(-2.43)	(-2.09)	(-0.09)	(-2.87)	(-2.44)	(-2.10)	(-0.12)
$Turn$	51.980***	46.678***	98.433***	92.495***	50.810***	45.541***	97.268***	90.962***	49.611***	44.214***	95.602***	88.938***
	(4.24)	(3.84)	(4.86)	(4.75)	(4.21)	(3.79)	(4.84)	(4.70)	(4.17)	(3.72)	(4.80)	(4.64)
$Size$	6.028***	5.968***	10.985***	11.340***	6.038***	5.980***	11.005***	11.361***	6.024***	5.964***	10.983***	11.334***
	(17.08)	(16.60)	(24.14)	(24.38)	(17.01)	(16.53)	(24.14)	(24.37)	(16.88)	(16.37)	(23.91)	(24.06)
Constant	-2.921	-2.380	-86.023***	-94.014***	-3.132	-2.610	-86.310***	-94.355***	-2.770	-2.208	-85.792***	-93.761***
	(-0.52)	(-0.42)	(-11.88)	(-12.85)	(-0.56)	(-0.45)	(-11.89)	(-12.86)	(-0.49)	(-0.38)	(-11.76)	(-12.69)
$\bar{R}^2$	0.245	0.234	0.344	0.355	0.244	0.233	0.344	0.354	0.242	0.231	0.343	0.353
Nobs	94,829	94,829	93,874	93,874	94,829	94,829	93,874	93,874	94,829	94,829	93,874	93,874

Table 5
Firm Size, Public News Arrival and Correlation Breakdown

This table presents the regression results of several conditional correlation estimates between spot and the underlying futures on news arrival proxy, the interaction between news arrival proxy and firms size and other control variables including day dummies. The dependent variables, alternatively, are CC_{ask}, CC_{bid}, CC_{ask}^{BEKK}, and CC_{bid}^{BEKK}. CC_{ask} (CC_{bid}) is the conditional correlation between spot and futures ask (bid) quote log returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC_{ask}^{BEKK} (CC_{bid}^{BEKK}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. σ_{CSS} is equal to the standard deviation of composite sentiment scores available over the day. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. $Large$ is a dummy variable which equals one if the stock belongs to the top size quartile, and zero otherwise. t-statistics shown in the parentheses are based on robust standard errors adjusted for both time clustering (by day) and firm-level clustering. ***, **, and * indicate 1%, 5% and 10% statistical significance, respectively. All the reported coefficients are scaled upward by 100. Nobs is the number of stock-day observations. The sample period is from 20 July 2009 to 19 March 2010.

Panel A: Public News Arrival Proxy $\#News2$

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#News2$	1.037***	0.939***	2.539***	2.718***	0.175***	0.134**	0.393***	0.377***
	(5.32)	(4.22)	(6.97)	(6.12)	(3.61)	(2.55)	(3.54)	(3.31)
$\#News2\times Size$	-0.071***	-0.067***	-0.168***	-0.182***				
	(-6.11)	(-5.10)	(-7.70)	(-6.58)				
$\#News2\times Large$					-0.233***	-0.205***	-0.466***	-0.491***
					(-4.41)	(-3.53)	(-3.91)	(-3.66)
$Price$	-0.039***	-0.033***	-0.011	0.014	0.062***	0.071***	0.187***	0.222***
	(-3.41)	(-3.08)	(-0.85)	(0.82)	(3.19)	(3.37)	(3.53)	(3.63)
$Turn$	79.352***	73.429***	167.165***	156.489***	77.076***	70.338***	156.099***	144.074***
	(5.29)	(5.03)	(5.90)	(5.80)	(4.47)	(4.24)	(4.75)	(4.60)
$Size$	7.253***	7.301***	14.305***	14.789***				
	(23.33)	(23.37)	(40.10)	(41.79)				
$Large$					16.050***	15.643***	29.548***	29.782***
					(17.53)	(16.63)	(18.38)	(16.91)
Constant	-22.846***	-23.832***	-140.658***	-150.369***	76.126***	75.857***	55.594***	52.621***
	(-4.59)	(-4.77)	(-25.58)	(-28.32)	(70.71)	(69.57)	(30.50)	(27.13)
$\bar{R}^2$	0.273	0.269	0.443	0.466	0.121	0.115	0.202	0.213
Nobs	205,603	205,603	201,683	201,683	205,603	205,603	201,683	201,683

Panel B: News Uncertainty Proxy σ_{CSS}

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
σ_{CSS}	1.575***	1.511***	3.744***	4.059***	-0.346***	-0.382***	-0.428**	-0.506***
	(2.58)	(2.66)	(4.01)	(3.66)	(-3.06)	(-3.53)	(-2.46)	(-2.90)
$\sigma_{CSS}\times Size$	-0.129***	-0.127***	-0.282***	-0.309***				
	(-3.27)	(-3.43)	(-4.67)	(-4.21)				
$\sigma_{CSS}\times Large$					0.172	0.163	0.241	0.179
					(1.30)	(1.26)	(1.21)	(0.75)
$Price$	-0.034***	-0.026**	-0.024**	-0.001	0.015	0.022**	0.066**	0.092***
	(-2.90)	(-2.45)	(-2.00)	(-0.05)	(1.48)	(2.00)	(2.56)	(2.70)
$Turn$	46.622***	40.340***	88.948***	81.351***	23.042**	16.625	40.667**	31.400*
	(3.94)	(3.41)	(4.53)	(4.24)	(2.09)	(1.58)	(2.23)	(1.86)
$Size$	6.249***	6.148***	11.538***	11.924***				
	(17.30)	(16.79)	(24.78)	(25.82)				
$Large$					11.180***	10.822***	19.687***	20.329***
					(13.99)	(13.10)	(15.17)	(14.54)
Constant	-6.420	-5.315	-94.657***	-103.167***	83.946***	83.643***	72.743***	69.835***
	(-1.12)	(-0.91)	(-12.74)	(-14.17)	(99.61)	(96.70)	(50.33)	(45.28)
$\bar{R}^2$	0.242	0.229	0.342	0.351	0.103	0.098	0.146	0.157
Nobs	94,829	94,829	93,874	93,874	94,829	94,829	93,874	93,874

Table 6
Volatile Market, Public News Arrival and Correlation Breakdown

This table presents the regression results of several conditional correlation estimates between spot and the underlying futures on the news arrival proxy, extreme market condition variable (denoted as $MVol$), the interaction term between news arrival proxy, firm size and extreme market condition variable. All other control variables including day dummies are also included. Conditional Correlation Proxies, alternatively, are CC_{ask}, CC_{bid}, CC_{ask}^{BEKK}, and CC_{bid}^{BEKK}. CC_{ask} (CC_{bid}) is the conditional correlation between spot and futures ask (bid) quote log returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC_{ask}^{BEKK} (CC_{bid}^{BEKK}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. σ_{CSS} is equal to the standard deviation of composite sentiment scores available over the day. $MVol$ is the absolute percentage return of the S&P500 Index. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. t-statistics shown in the parentheses are based on robust standard errors adjusted for both time clustering (by day) and firm-level clustering. ***, **, and * indicate 1%, 5% and 10% statistical significance, respectively. All the reported coefficients are scaled upward by 100. Nobs is the number of stock-day observations. The sample period is from 20 July 2009 to 19 March 2010.

Panel A: Public News Arrival Proxy $\#News2$

	CC_{ask}		CC_{bid}		CC_{ask}^{BEKK}		CC_{bid}^{BEKK}	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#News2$	0.992***	1.278***	0.888***	1.163***	2.517***	2.830***	2.701***	2.940***
	(5.03)	(5.54)	(3.91)	(4.12)	(6.93)	(6.95)	(6.09)	(5.93)
$\#News2\times MVol$	0.059**	-0.310**	0.067***	-0.289*	0.030	-0.376**	0.022	-0.288
	(2.52)	(-2.38)	(3.04)	(-1.77)	(1.31)	(-1.96)	(1.04)	(-1.42)
$\#News2\times Size$	-0.071***	-0.088***	-0.067***	-0.084***	-0.168***	-0.187***	-0.182***	-0.196***
	(-6.05)	(-6.17)	(-5.03)	(-4.88)	(-7.69)	(-7.48)	(-6.58)	(-6.33)
$\#News2\times Size\times MVol$		0.022***		0.021**		0.024**		0.019
		(2.60)		(2.05)		(2.02)		(1.48)
$MVol$	-2.637***	-2.596***	-2.852***	-2.813***	-1.720**	-1.676**	-1.325**	-1.291**
	(-3.10)	(-3.09)	(-3.56)	(-3.56)	(-2.25)	(-2.22)	(-2.07)	(-2.04)
$Price$	-0.039***	-0.039***	-0.034***	-0.034***	-0.012	-0.012	0.014	0.014
	(-3.43)	(-3.43)	(-3.11)	(-3.11)	(-0.88)	(-0.88)	(0.81)	(0.81)
$Turn$	81.465***	81.549***	75.699***	75.780***	168.747***	168.836***	157.710***	157.778***
	(5.31)	(5.32)	(5.05)	(5.06)	(5.89)	(5.90)	(5.79)	(5.80)
$Size$	7.256***	7.257***	7.304***	7.305***	14.310***	14.310***	14.792***	14.793***
	(23.37)	(23.39)	(23.42)	(23.43)	(40.15)	(40.17)	(41.83)	(41.84)
Constant	-20.923***	-20.964***	-21.752***	-21.792***	-139.440***	-139.485***	-149.431***	-149.465***
	(-4.15)	(-4.16)	(-4.33)	(-4.34)	(-25.27)	(-25.30)	(-27.89)	(-27.91)
$\bar{R}^2$	0.278	0.278	0.275	0.275	0.444	0.444	0.466	0.466
Nobs	205,603	205,603	205,603	205,603	201,683	201,683	201,683	201,683

Table 6 - Continued

Panel B: News Uncertainty Proxy σ_{CSS}

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
σ_{CSS}	1.710***	3.534***	1.597***	3.444***	3.949***	5.705***	4.246***	5.502***
	(2.78)	(4.07)	(2.80)	(4.22)	(4.24)	(5.35)	(3.82)	(4.43)
$\sigma_{CSS}\times MVol$	-0.096**	-2.311***	-0.048	-2.291***	-0.164***	-2.273***	-0.156***	-1.662***
	(-2.19)	(-3.14)	(-1.15)	(-3.36)	(-3.14)	(-3.44)	(-2.68)	(-2.74)
$\sigma_{CSS}\times Size$	-0.132***	-0.253***	-0.130***	-0.252***	-0.286***	-0.402***	-0.313***	-0.396***
	(-3.36)	(-4.56)	(-3.50)	(-4.79)	(-4.76)	(-5.85)	(-4.27)	(-4.89)
$\sigma_{CSS}\times Size\times MVol$		0.147***		0.149***		0.140***		0.100***
		(3.07)		(3.34)		(3.21)		(2.60)
$MVol$	-1.004**	-1.169**	-1.253**	-1.419**	-0.415	-0.565	-0.188	-0.295
	(-1.96)	(-2.09)	(-2.45)	(-2.56)	(-0.69)	(-0.89)	(-0.35)	(-0.53)
$Price$	-0.034***	-0.034***	-0.027**	-0.027**	-0.024**	-0.024**	-0.001	-0.001
	(-2.91)	(-2.91)	(-2.47)	(-2.47)	(-2.02)	(-2.02)	(-0.06)	(-0.06)
$Turn$	47.351***	47.246***	41.146***	41.040***	89.509***	89.432***	81.747***	81.692***
	(3.95)	(3.95)	(3.43)	(3.43)	(4.53)	(4.53)	(4.23)	(4.23)
$Size$	6.264***	6.280***	6.162***	6.179***	11.553***	11.572***	11.936***	11.949***
	(17.34)	(17.35)	(16.83)	(16.84)	(24.80)	(24.79)	(25.85)	(25.85)
Constant	-5.913	-6.051	-4.612	-4.752	-94.592***	-94.776***	-103.224***	-103.356***
	(-1.02)	(-1.04)	(-0.79)	(-0.81)	(-12.71)	(-12.72)	(-14.14)	(-14.15)
$\bar{R}^2$	0.244	0.245	0.231	0.233	0.343	0.343	0.351	0.352
Nobs	94,829	94,829	94,829	94,829	93,874	93,874	93,874	93,874

Table 7
Robustness Tests

This table presents two sets of robustness tests on the variations of the regression results of conditional correlation estimates between spot and the underlying futures on public news arrival proxies. Panel A replicates the previous analyses using search frequency data from Google (denoted as SVI) as the news arrival measure. Panels B and C repeat the analyses using Tobit regression methods. Conditional correlation proxies, alternatively, are CC_{ask}, CC_{bid}, CC^{BEKK}_{ask}, and CC^{BEKK}_{bid}. CC_{ask} (CC_{bid}) is the conditional correlation between spot and futures ask (bid) quote log returns at 10-minute intervals in a forward rolling window period over six 10-minute intervals. CC^{BEKK}_{ask} (CC^{BEKK}_{bid}) is the conditional correlation between the spot ask (bid) and futures ask (bid) returns derived from the VEC-BEKK model. $\#News2$ is the sum of all available news relevance scores scaled by 100 on a given trading day. σ_{CSS} is equal to the standard deviation of composite sentiment scores available over the day. $MVol$ is the absolute percentage return of the S&P500 Index. $Price$ is the closing share price of the underlying stock. $Turn$ is the number of shares of underlying stocks traded scaled by the number of shares outstanding. $Size$ is the natural log of market capitalization in thousands. t-statistics shown in the parentheses are based on robust standard errors adjusted for both time clustering (by day) and firm-level clustering. ***, **, and * indicate 1%, 5% and 10% statistical significance, respectively. All the reported coefficients are scaled upward by 100. Nobs is the number of stock-day observations. The sample period is from 20 July 2009 to 19 March 2010.

Panel A: An Alternative Measure of News Arrival

	CC_{ask}	CC_{bid}	CC^{BEKK}_{ask}	CC^{BEKK}_{bid}	CC_{ask}	CC_{bid}	CC^{BEKK}_{ask}	CC^{BEKK}_{bid}	CC_{ask}	CC_{bid}	CC^{BEKK}_{ask}	CC^{BEKK}_{bid}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
SVI	-0.507***	-0.515***	-0.645*	-0.448	-2.553	-2.014	6.942**	12.697***	7.586**	8.355***	13.125***	17.449***
	(-2.73)	(-2.79)	(-1.94)	(-1.34)	(-1.45)	(-1.19)	(2.38)	(3.75)	(2.22)	(2.76)	(3.13)	(3.81)
$SVI\times Size$					0.140	0.103	-0.519***	-0.900***	-0.585**	-0.638***	-0.958***	-1.233***
					(1.24)	(0.94)	(-2.74)	(-4.06)	(-2.44)	(-3.00)	(-3.44)	(-4.11)
$SVI\times MVol$									-13.954***	-14.253***	-8.469***	-6.470**
									(-3.60)	(-3.92)	(-2.87)	(-2.36)
$SVI\times Size\times MVol$									1.005***	1.025***	0.605***	0.458**
									(3.53)	(3.83)	(2.91)	(2.47)
$MVol$									-5.972***	-5.764***	-3.868***	-2.989***
									(-3.93)	(-3.84)	(-3.51)	(-3.16)
$Price$	-0.053***	-0.043***	-0.021	0.007	-0.053***	-0.043***	-0.021	0.008	-0.055***	-0.045***	-0.022	0.007
	(-3.90)	(-3.54)	(-1.56)	(0.37)	(-3.91)	(-3.55)	(-1.55)	(0.42)	(-3.96)	(-3.62)	(-1.63)	(0.38)
$Turn$	91.044***	86.852***	195.035***	192.594***	91.769***	87.384***	192.345***	187.933***	95.371***	90.769***	194.742***	189.793***
	(4.55)	(4.33)	(6.89)	(7.55)	(4.60)	(4.37)	(6.78)	(7.31)	(4.61)	(4.38)	(6.76)	(7.27)
$Size$	6.290***	6.096***	12.749***	13.610***	6.196***	6.027***	13.099***	14.217***	6.185***	6.016***	13.093***	14.213***
	(12.65)	(12.34)	(31.19)	(33.13)	(12.22)	(11.98)	(32.18)	(34.85)	(12.24)	(11.98)	(32.23)	(34.73)
Constant	-6.677	-4.144	-116.066***	-131.954***	-5.345	-3.168	-121.006***	-140.512***	-0.966	1.054	-118.184***	-138.344***
	(-0.81)	(-0.50)	(-17.85)	(-20.82)	(-0.64)	(-0.38)	(-18.69)	(-22.38)	(-0.12)	(0.13)	(-18.46)	(-21.80)
$\bar{R}^2$	0.205	0.200	0.416	0.455	0.205	0.200	0.417	0.456	0.230	0.225	0.421	0.459
Nobs	74,797	74,797	74,706	74,706	74,797	74,797	74,706	74,706	74,797	74,797	74,706	74,706

Table 7 - Continued

Panel B: Tobit model using news arrival proxy

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\#News2$	-0.142***	-0.173***	-0.249***	-0.301***	1.037***	0.939***	2.539***	2.718***	1.278***	1.163***	2.830***	2.940***
	(-4.56)	(-5.11)	(-4.72)	(-4.52)	(5.32)	(4.22)	(6.97)	(6.12)	(5.54)	(4.12)	(6.95)	(5.93)
$\#News2 \times Size$					-0.071***	-0.067***	-0.168***	-0.182***	-0.088***	-0.084***	-0.187***	-0.196***
					(-6.11)	(-5.10)	(-7.70)	(-6.58)	(-6.17)	(-4.88)	(-7.48)	(-6.33)
$\#News2 \times MVol$									-0.310**	-0.289*	-0.376**	-0.288
									(-2.38)	(-1.77)	(-1.96)	(-1.42)
$\#News2 \times Size \times MVol$									0.022***	0.021**	0.024**	0.019
									(2.60)	(2.05)	(2.02)	(1.48)
$MVol$									-2.596***	-2.813***	-1.676**	-1.291**
									(-3.09)	(-3.56)	(-2.22)	(-2.04)
$Price$	-0.040***	-0.034***	-0.013	0.012	-0.039***	-0.033***	-0.011	0.014	-0.039***	-0.034***	-0.012	0.014
	(-3.59)	(-3.22)	(-0.85)	(0.63)	(-3.41)	(-3.08)	(-0.85)	(0.82)	(-3.43)	(-3.11)	(-0.88)	(0.81)
$Turn$	86.579***	80.240***	184.232***	174.962***	79.352***	73.429***	167.165***	156.489***	81.549***	75.780***	168.836***	157.778***
	(5.71)	(5.43)	(6.42)	(6.39)	(5.29)	(5.03)	(5.90)	(5.80)	(5.32)	(5.06)	(5.90)	(5.80)
$Size$	7.108***	7.164***	13.951***	14.406***	7.253***	7.301***	14.305***	14.789***	7.257***	7.305***	14.310***	14.793***
	(23.11)	(23.06)	(38.40)	(39.27)	(23.33)	(23.37)	(40.10)	(41.79)	(23.39)	(23.43)	(40.17)	(41.84)
Constant	-20.702***	-21.812***	-135.438***	-144.719***	-22.846***	-23.832***	-140.658***	-150.369***	-20.964***	-21.792***	-139.485***	-149.465***
	(-4.21)	(-4.39)	(-24.34)	(-26.55)	(-4.59)	(-4.77)	(-25.58)	(-28.32)	(-4.16)	(-4.34)	(-25.30)	(-27.91)
Nobs	205,603	205,603	201,683	201,683	205,603	205,603	201,683	201,683	205,603	205,603	201,683	201,683

Panel C: Tobit model using news uncertainty proxy

	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}	CC_{ask}	CC_{bid}	CC_{ask}^{BEKK}	CC_{bid}^{BEKK}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\#News2$	-0.384***	-0.423***	-0.541***	-0.649***	1.575***	1.511***	3.744***	4.059***	3.534***	3.444***	5.705***	5.502***
	(-5.75)	(-6.51)	(-5.28)	(-5.79)	(2.58)	(2.66)	(4.01)	(3.66)	(4.07)	(4.22)	(5.35)	(4.43)
$\#News2 \times Size$					-0.129***	-0.127***	-0.282***	-0.309***	-0.253***	-0.252***	-0.402***	-0.396***
					(-3.27)	(-3.43)	(-4.67)	(-4.21)	(-4.56)	(-4.79)	(-5.85)	(-4.89)
$\#News2 \times MVol$									-2.311***	-2.291***	-2.273***	-1.662***
									(-3.14)	(-3.36)	(-3.44)	(-2.74)
$\#News2 \times Size \times MVol$									0.147***	0.149***	0.140***	0.100***
									(3.07)	(3.34)	(3.21)	(2.60)
$MVol$									-1.169**	-1.419**	-0.565	-0.295
									(-2.09)	(-2.56)	(-0.89)	(-0.53)
$Price$	-0.034***	-0.027**	-0.024*	-0.001	-0.034***	-0.026**	-0.024**	-0.001	-0.034***	-0.027**	-0.024**	-0.001
	(-2.92)	(-2.46)	(-1.93)	(-0.07)	(-2.90)	(-2.45)	(-2.00)	(-0.05)	(-2.91)	(-2.47)	(-2.02)	(-0.06)
$Turn$	47.239***	40.949***	90.353***	82.895***	46.622***	40.340***	88.948***	81.351***	47.246***	41.040***	89.432***	81.692***
	(3.97)	(3.44)	(4.57)	(4.29)	(3.94)	(3.41)	(4.53)	(4.24)	(3.95)	(3.43)	(4.53)	(4.23)
$Size$	5.869***	5.772***	10.711***	11.015***	6.249***	6.148***	11.538***	11.924***	6.280***	6.179***	11.572***	11.949***
	(16.51)	(15.89)	(23.09)	(22.44)	(17.30)	(16.79)	(24.78)	(25.82)	(17.35)	(16.84)	(24.79)	(25.85)
Constant	-0.663	0.366	-82.111***	-89.386***	-6.420	-5.315	-94.657***	-103.167***	-6.051	-4.752	-94.776***	-103.356***
	(-0.12)	(0.06)	(-11.14)	(-11.64)	(-1.12)	(-0.91)	(-12.74)	(-14.17)	(-1.04)	(-0.81)	(-12.72)	(-14.15)
Nobs	94,829	94,829	93,874	93,874	94,829	94,829	93,874	93,874	94,829	94,829	93,874	93,874