Non-scheduled news arrival and high-frequency stock market dynamics

Evidence from the Australian Securities Exchange

An increasing number of market participants utilize news analytics software to comprehend the large amounts of unstructured data flowing through news-wires. Utilizing original data from one such tool – Ravenpack – I examine the market reaction of leading Australian stocks to stock-specific news flow over an extended period. Unconditional analysis of key variables around 484,440 news items reveals distinct responses in market activity, volatility, bid-ask spreads and returns. The study confirms previous literature such that indicated relevance of news items is critical when identifying significant effects. In addition, the reaction of market activity, volatility and spreads is greatest for negative news. The findings are confirmed when controlling for market dynamics and cross-dependencies between variables in a high-frequency VAR model.

Key Words: News analytics; Stock market; Non-scheduled news events
JEL Classification Codes: G02, G14, G15

1. Introduction

Market efficiency suggests that all currently available public, and private, information should be reflected in share prices. Market participants should only respond to new information (news), and therefore price movements and trading activity will be strongly influenced by the released of both scheduled and non-scheduled news. However, with the advent of modern technology the flow of news has greatly increased and this makes it costly for market participants to process all asset-specific news. As a result, many participants are starting to rely on pre-processed news analytics provided by news vendors; with such data playing an increasingly important role in the trading of financial assets it seems pertinent to ask whether the indicators of relevance and sentiment are both useful and reliable.

Historically, research in this field has focused on specific and readily quantifiable news events such as scheduled macroeconomic announcements and earnings results. For example, Patell and Wolfson (1984) and Woodruff and Senchak (1988) consider the adjustment of stock prices following earnings and dividend announcements, and find that much of the market adjustment occurs in the first 30 minutes following the announcement. Ederington and Lee (1993, 1995), Becker et al. (1996), Bernanke and Kuttner (2005), Rigobon and Sack (2004), and Smales (2013) are among the many papers that consider the impact of macroeconomic announcements with the confirmation of a dynamic response to the news (surprise) component of data releases that quickly subsides.

More recently, the quantifying of news language, by researchers such as Antweiler and Frank (2004) and Tetlock (2007), has enabled the identification of common patterns in firm responses and market reactions across a wider range of events. In particular, the relevance and sentiment of news has been tested in a variety of market settings. Tetlock el al. (2008) examine whether a quantitative measure of language can be used to predict firms' earnings and stock returns, and find that negative words in firm-specific news stories forecast low firm earnings. Sinha (2011) gauges the tone of news articles and constructs a measure to predict future returns while Engelberg et al. (2012) find that the negative relation between short sales and future returns is significantly stronger in the presence of news stories containing negative news. Dzielinski (2011) utilizes sentiment signed news to directly compare news and no-news stock returns, finding that positive (negative) news results in above (below) average returns whilst the effect of neutral news is non-distinguishable from the no-news average. Interestingly, Tetlock (2011) also reports that investors react to stale news.

In terms of framework, this paper is most similar to Groß-Klußmann and Hautsch (2011) who examine high-frequency news-implied market reactions on 39 liquid stocks traded on the London Stock Exchange over an 18 month period from January 2007. They observe that trading activity reacts significantly to company-specific news items that are identified as relevant, although they do not consider the importance of directional sentiment indicators.

In addition to examining the interdependency of a number of key market variables, this paper primarily asks a single key question: do indicators of relevance and sentiment matter, and if so, does pre-processed data do a good job of assigning such indicators? The paper develops the existing literature in two important ways. Firstly, examination of the Australian market allows the consideration of whether the existing findings are applicable in a broader international context. Secondly, this is the first study to examine the impact of non-scheduled news flow on market activity over such an extended time period; a period which neatly encapsulates the global financial crisis (GFC) of 2007-2009 and thus allows for an initial examination of the impact of crisis on market efficiency surrounding non-scheduled news announcements. Importantly, the ASX sample differs from that of many major markets in two ways. First, during the sample period pre-dates the arrival of Chi-X into the Australian market, and as such there was no rival exchange for trading Australian shares; hence all trading activity surrounding the non-scheduled announcements is captured using the ASX data-set. Second, the ASX is dominated by just two industries with financial services (44.8%) and materials (17.4%) constituting nearly a two-thirds of total market capitalisation as of December 2012[1].

Using a sample of 33 highly-liquid ASX50 stocks over a 12-year time period, high-frequency (30 sec interval) market activity around 484,440 non-scheduled news announcements is examined. High-relevance news items induce an increase in market activity, volatility and spreads, whilst news with negative sentiment has a greater impact than positive news, and the

[1] This compares to the S&P500 where the two largest industries (I.T. and financials) make-up 34% of the total index, and the FTSE100 where the two largest (Financials and Oil & Gas) total less than 40% of the index.

impact of neutral news is indistinguishable from no-news. There is evidence that returns react to news prior to its arrival on the news-wire but the level of returns is not sufficient to be profitable after considering transaction costs. Analysis of market dynamics and cross-dependencies between variables in a VAR framework confirms the significant market impact induced by contemporaneous news items and also reveals a significant and positive relationship between the measures of trading activity and volatility. Consistent with established microstructure theory, there is an increase in bid-ask spreads if prior trading periods reflect rising market activity and volatility. In addition, an event study analysis of the largest Australian banking stocks reveals that the impact of news on market activity and volatility is greater following the onset of the Global Financial Crisis (GFC) in 2007. Indeed, for returns the relationship with news releases is non-existent prior to the GFC, but significant during the GFC.

The remainder of this paper is organized as follows: Section 2 discusses the nature of the data used in this paper, with a particular focus on the news analytics tool utilised. Section 3 investigates the unconditional effect of news items on stock prices after disaggregating news by relevance and sentiment. Section 4 introduces a high-frequency VAR model in order to control for market dynamics and cross-dependencies. Section 5 concludes the paper.

2. Data

Several information vendors offer software that captures the important characteristics in real-time[2]. The software uses pattern recognition algorithms to analyse the text of news releases to infer tone and sentiment, as well as relative importance. In this study, processed news data is gathered from a software tool named Ravenpack[3]; this package utilises news items posted on the Dow Jones newswire and in the Wall Street Journal. In total, there are 484,400 news headlines for the stocks I consider over the period 04 Jan 2000 to 01 Nov 2011, with news arrival recorded with millisecond precision. Ravenpack provides a number of indicators for each news message although the focus here is on Relevance, Novelty, and Sentiment.

Relevance is given by a number in the [0,100] interval, indicating how strongly related the company is to the underlying story, with higher values indicating greater relevance. For any

[2] This processed data is available to market participants (at a cost) almost instantaneously; academic researchers are able to access this information only at a later stage – usually several months afterwards.
[3] More extensive information on the nature of the Ravenpack news analytics tool may be found at www.ravenpack.com

news story that mentions a company, Ravenpack provides a relevance score. A score of 0 means the company was passively mentioned while a score of 100 means the company was predominant in the news story. This analysis defines an event with low relevance as having a relevance score below 10, whilst high relevance has a relevance score above 90. Ex-ante expectations are for news with high-relevance to have a greater market impact. The *Novelty* of a news item is measured in Ravenpack using the RP_Story_Event_Index; a new story that is published for the first time has an index value equal to 1. In order to ensure analysis focuses on the arrival of *new* information items with an Event_Index greater than 1 are excluded from the sample.

<Insert Figure 1>

Ravenpack provides several *Sentiment* indicators with *Multi-Classifier for Equities* (MCQ) providing the focus in this paper. MCQ is a multi-classifier score that represents the news sentiment on the tone applicable only towards the most relevant companies mentioned in a story. The score is derived from a combination of analytics values produced by classifiers which focus on short commentary and editorials (BMQ), earnings evaluations (BEE), corporate actions (BCA) and changes in analyst recommendations (ANL_CHG). BEE uses Ravenpack's Traditional Tagging methodology and is based on a Rule Base that maps key words, phrases, combinations and other word-level definitions to pre-defined sentiment values. An Expert Consensus Methodology underpins BMQ and BCA scores and entails training classification algorithms on the results of financial experts manually tagging stories. An MCQ score is present when the relevance score for a company is 90 or higher and either there is an ANL-CHG score or all of BMQ, BEE and BCA scores are positive, neutral or negative. The logic of this classifier is to detect consistent sentiment classifications, and discard combinations where the classifiers have contradictory scores. Ravenpack assigns this classifier a score of 0 to negative sentiment, 50 to neutral and 100 to positive; the analysis scales the scores to the more intuitive levels of -1, 0, +1.

As the aim of this paper is to focus on unscheduled intraday news driven mostly by random events, I specifically exclude the BER classifier which focuses on scheduled earnings releases. In addition, only news flow which occurs within the trading day, which runs from 10:00 to 16:00 (AEST) is considered.

Figure 1 provides a graphical representation of the news variables considered in this paper. The distribution of relevance and MCQ indicators are depicted in Fig. 1a; approximately

40% of news items have a relevance score greater than 0.9 and are thus considered highly relevant, of those highly relevant items 6.85% are deemed negative and 13.22% positive. The first panel in Fig. 1b shows that news arrival is highest at the start of the trading day (10:00am) and gradually falls to an equilibrium level over the course of the morning. Spikes in news arrival occur every hour on the hour, this is possibly due to regular hourly news bulletins which will cover a number of different companies and cover the largest companies / most liquid stocks (which are the same stocks as covered in the sample). On average 0.22 news items arrive in each 30 second interval (alternatively the average arrival rate of news items is every 136 secs). The second panel of Fig. 1b shows the distribution of news items per trading day over the sample period of 04 Jan 2000 to 01 Nov 2011. On average there are approximately 320 news items reported in each trading day and this has remained relatively constant over time; whilst there is no pronounced yearly pattern of news arrival there are noticeable spikes which have occurred around specific events such as the terrorist attacks on 11 Sep 2001, and the demise of Lehman Brothers in 2008.

To enable a high-frequency study of market dynamics it is necessary to ensure that the stocks covered within the analysis are liquid and thus I focus on the stocks which make up the ASX50; the leading 50 domestic stocks by market capitalisation trading on the Australian Securities Exchange (ASX). As data is required for the 2,940 trading days that make up the Jan '00 to Nov '11 sample, the sample is reduced to the 33 stocks listed for the entire sample period. Covering 75.7% of the market capitalization of the Australian All Ordinaries Index, the sample can be considered as being representative of the Australian stock market[4] (Table 1).

<Insert Table 1>

Underlying transaction data is obtained from Thomson Reuters Tick History, via SIRCA[5], and is aggregated into 30 second intervals. This aggregation level is a compromise between exploiting maximum information and making the analysis computationally tractable and this is especially important given that the analysis covers 12 years of data. Market activity, volatility and liquidity are captured by the following variables computed over 30 second intervals.

i) Money value traded, calculated as price multiplied by volume traded in the given interval;

[4] In terms of the number of stocks, and percentage of total market capitalisation, the sample is similar in nature to the LSE sample utilised by Groß-Klußmann and Hautsch (2011)

[5] Securities Industry Research Centre of Asia-Pacific

ii) Volatility, calculated using volatility of mid-point returns in each interval;

iii) Absolute trade imbalance, defined as the absolute value of the difference in cumulated buyer- and seller- initiated trades (identified with the Lee and Ready (1991) algorithm);

iv) Average trade size, defined as total volume divided by the corresponding number of trades per interval;

v) Bid-ask spread, defined as the average bid-ask spread over the given interval;

vi) Returns, calculated using the mid-point of the bid-ask quote;

Following Groß-Klußmann and Hautsch (2011), each of the variables are standardized, by the yearly average of the corresponding underlying 30-second interval, to account for pronounced intraday patterns. Thus, standardized variables are computed according to:

$$x_{jd}^{*} = \frac{x_{jd}}{1/n \sum_{d=1}^{n} x_{jd}} \qquad (1)$$

Where j denotes the specific interval of the trading day d and x represents the corresponding variable[6,7].

3. Unconditional Effects of News Items

3.1 Impact of high relevance news on volatility and liquidity

I first gain an insight into the impact of news on market activity by quantifying the unconditional impact of high relevance news without controlling for market dynamics and cross-dependencies between variables. I analyse 90 30-second intervals around the arrival of news items capturing 30 intervals (15-minutes) before each disclosure and 60 intervals (30-minutes) afterwards.

<Insert Figure 2>

The timing of the intervals is illustrated in figure 2. I_0 denotes the specific 30-second interval starting at the release of the news item. For each stock, the average market reaction and

[6] The standardized variables are tested for unit roots using the Augmented Dickey-Fuller method; the presence of a unit root was rejected at the 1% level for all variables.

[7] The Bai-Perron (2003) is used to test for structural breaks in the data and inconclusive evidence is found for the majority of variables. The exception are the return and volatility variables for the major banks starting with the onset of the global financial crisis in 2007; this possible structural break is addressed as part of the case study in Section 3.4.

corresponding standard errors are computed over all event windows. For the purpose of conciseness, only pooled averages over the cross-section of stocks are reported. Consequently, by denoting the market reaction of variable X to news item i during interval I_j as X_{iI_j}, the pooled average across all news events and all stocks is computed as $\bar{X}_{I_j} = 1/n \sum_{i=1}^{n} X_{iI_j}$, where n is the total number of news items for all stocks. Given that the stocks have quite similar empirical characteristics this process allows one to highlight the results common to all stocks. Figure 3 shows the dollar-volume of shares traded, realized volatility, bid-ask spreads, average trade sizes and absolute trade imbalances around high relevance and low-relevance news items. Note that by construction of the seasonality adjustment the mean of each series equals one.

<Insert Figure 3>

The following findings can be summarized: First, I identify significant upward movements in volume of shares traded and volatility around the releases of news items (Fig. 3a.); that is, news releases induce increases in trading activity and corresponding increases in volatility. Easley and O'Hara (1987) suggest larger trade sizes (Fig. 3b.) are due to execution by better informed market participants, while Harris and Raviv (1993) suggest that trading occurs due to differences in opinion of traders on news' topics. Increases in absolute trade imbalance (Fig. 3b.) suggest that trading activity becomes more asymmetric in periods of information dissemination. Second, the release of a high-relevance news item significantly increases bid-ask spreads compared to low-relevance news (Fig. 3c.), indeed spreads widen further as the news-release of approaches; liquidity suppliers react to news by reducing order aggressiveness in revising quotes to avoid the costs incurred with adverse selection by trading with informed traders.

Third, the Ravenpack-indicated relevance of a news item is clearly supported by corresponding market reactions. All variables maintain a higher level, and respond significantly stronger to the news, if information is indicated to be of high-relevance. For low-relevance news if is not possible to identify significant deviations of the analysed trading variables from their pre-news levels. Fourth, all variables exhibit above-average levels starting more than 15-minutes before the arrival of scheduled news-items – a phenomenon which also occurs in the case of periodically scheduled earnings releases. This result indicates that some market participants seem to have additional and timelier channels of information. It may also be possible to attribute the pre-news reaction to a clustering of news, although by limiting the Event_Index to a score of not greater than 1, i.e. novel news stories only, it is hoped to reduce this impact in the analysis.

3.2 Relative impact of negative and positive news items

Fig.4. depicts the volume of shares traded, realized volatility, bid-ask spreads, average trade sizes and absolute trade imbalances around high relevance news items which have been disaggregated into negative, neutral and positive sentiments using Ravenpack's MCQ variable.

<Insert Figure 4>

Again, it is possible to summarize the findings across variables. Consistent with Dzielinski (2011), neutral news releases appears to result in no significant change in any of the variables around the time of the news release. In contrast, both positive and negative news releases elicit a sharp spike in the normalized variables in the period around news releases. The above-average activities start at least 15-minutes prior to the news release, move sharply higher in the 2-minutes immediately prior to the release, and quickly revert to lower levels following the release.

In addition, the measures for all normalized variables are higher in the case of negative news, particular in the 30 second intervals either side of the news release; trading activity and volatility increases despite wider than average spreads. Such results are consistent with findings by De Bondt and Thaler (1985), and Chen et al (2003), such that negative news induces a more significant market reaction.

3.3 Impact on returns

To test for the profitability of trading on news items I employ an event study framework as outlined in Campbell et al. (1997), and utilized by Groß-Klußmann and Hautsch (2011). As a model for 'normal' returns the following market model is assumed:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \gamma_i R_{i,t-1} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \left(0, \sigma_i^2\right) \qquad (2)$$

Where t denotes the underlying (30s) intervals, R_{mt} is the market return, computed as the return of the ASX 200 Index, and R_{it} is the return for stock i. To capture return dynamics on high frequencies I also include lagged returns. Using the resulting parameter estimates, I compute the abnormal returns $\widehat{AR}_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt} - \hat{\gamma}_i R_{i,t-1}$during the event windows. I then calculate the cumulative abnormal return in the normal manner and average over all stocks.

<Insert Figure 5>

Figure 5 shows the averaged cumulated abnormal returns around highly relevant news – disaggregated into positive and negative news events as defined by the MCQ measure. Starting 15 minutes *before* the disclosure I observe significantly positive (negative) cumulated abnormal returns as reactions to positive (negative) news items. However, whilst significant price movements are observed prior to news releases there are only limited return reactions thereafter. This evidence again points to the possibility of the presence of private pre-release information (i.e. information leakage) and clustering of news items. However, the magnitudes of returns are sufficiently small as to be non-profitable from a trading perspective once even moderate transaction costs are considered.

3.4 Event Study: The Big 4 Australian Banks

This section investigates the effect of the Global Financial Crisis (GFC) on the news reaction of the Big 4 Australian Banks (ANZ, CBA, NAB, WBC). This sub-section of firms is chosen on the basis that the nature of the GFC is such that the Australian banking sector was heavily affected as off-shore funding dried-up and as such the Big 4 stocks likely reacted strongly to news releases at this time. Additionally the 4 stocks are of interest as they are heavily traded and constitute 21.7% of total Australian market capitalisation. Finally, analysis of possible structural breaks in the time series of standardised variables suggests a possible break in the return, and volatility of return, measures for each of the banks in the period around July 2007[8]. The event study takes 17 July 2007 as the start of the GFC as this is when Bear Sterns first came public with news that problems existed with funds invested in mortgage backed securities, an announcement which precipitated the start of the crisis. The sample is therefore split into two sections; pre-GFC running from January 2000 to July 2007, and post-GFC which is deemed to cover the period July 2007 to December 2010. Since variables are normalized using annual data the reported variables should already account for any additional news impact following the GFC, so any differences that are reported will be stronger and economically significant.

[8] Analysis conducted using Bai and Perron (2003) method.

<Insert Figure 6>

Following Ederington and Lee (1993, 1995), and Smales (2013), Figure 6 depicts the disaggregated data adopting a tighter 12-minute window around the news release; with the window running from two minutes prior to the news arrival to 10-minutes after the news is published. In general there is little difference between the pre- and post-GFC results outside of the intervals immediately prior to and post the news release. However, all post-GFC variables have a greater normalized value, than pre-GFC variables, immediately prior to news arrival in the case of both negative and positive news; indicating a more vigorous reaction to news during the crisis. In addition, prior to the news release there is a significant difference in the level of market activity variables (money value traded, volatility, and absolute order imbalance) for negative and positive news releases in the post-GFC period; indicating that prior to negative news there is a stronger level of pre-emptive market activity for the Big 4 banks. This difference quickly dissipates following news arrival.

The pattern of post-crisis returns (Fig. 7), together with the increase in market activity immediately prior to news releases is again suggestive of the presence of informed traders or the clustering of news arrivals. However, evidence is more supportive of the likelihood that findings are driven by news clustering; returns do not exhibit the same pattern prior to the crisis, market-makers do not appear to adjust their spreads to reflect a greater chance of informed trades as Kyle (1985) would predict, and average trade size does not increase significantly as Easley and O'Hara (1987) suggest.

<Insert Figure 7>

4. Market Dynamics around news releases

4.1 VAR Model

The unconditional analysis in the previous section provides initial evidence of non-scheduled news invoking significant market reactions. However, apart from period returns, the variables exhibit significant autocorrelations (Fig. 8), indicating a high degree of persistence of the individual processes; consistent with Groß-Klußmann and Hautsch (2011). In addition, there appears to be significant cross-correlations between the variables.

<Insert Figure 8>

To correctly model the market dynamics around news releases the dependencies and interdependencies have to be explicitly taken into account. A six-dimensional model for the endogenous variables (realized variance, the money traded value, the bid-ask spread, average trade size, absolute trade imbalance, and period return) is utilized. The Vector Auto-Regression (VAR) specification is as follows:

$$y_t = c + \sum_{i=1}^{p} (\Gamma_i y_{t-i}) + \Xi \cdot D_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Omega) \tag{3}$$

Where Γ_i and Ξ denote (6 x 6) and (6 x (p_1+p_2+1)) coefficient matrices, and where $p_1 > 0$ and $p_2 > 0$ are integers. To capture the impact of news the dummy variable d_t is defined, taking a value one in case of relevant news in t and zero otherwise. Then, $D_t = (d_{t+p1} \ldots d_{t-p2})'$ is a vector of time dummies indicating the arrival of high relevance news and covering p_1 intervals before and p_2 intervals after news releases. The VAR model is applied to each of the 33 stocks in the sample. Akaike Information Criteria is utilised to obtain optimal lag length; in general the optimal lag length for each of the variables is found to be in the region of 6-12 lags and only the first three legs are reported in Table 2.

<Insert Table 2>

4.2 Estimation results

Table 2 reports average estimates for the VAR model augmented with dummies indicating the arrival of relevant news items. For conciseness, coefficients for lags of the dependent variables greater than three are not shown. News dummies cover the interval from 60 secs prior (t+2) to the news release to 120 secs after (t-4).

The results are summarized as follows: First, as expected from the analysis of the underlying autocorrelations, there is significant positive own dynamics for all variables. Second, high-frequency returns do not have a significant interaction with any of the other variables. Third, as one might expect, there is a significant positive relationship between the variables concerned with trading activity (Money Value Volume and Absolute Order Imbalance), and average trade size. Fourth, a significant and positive relationship is observed between the measures of trading activity and volatility, and this is particularly strong for money volume traded.

Consistent with asymmetric based microstructure theory (e.g. Easley and O'Hara (1992)), there is an increase in spreads if prior trading periods reflect rising market activity (measured by volume and order imbalance) and volatility. Since such situations are also characterized by declining trade size this is consistent with informed traders attempting to disguise their intentions when in possession of valuable news. In response to increased trading costs (indicated by higher bid-ask spreads) trading activity, or liquidity demand, is reduced.

Consistent with the unconditional analysis in Section 3, significant effects induced by contemporaneous news items are identified for all variables apart from returns. The relationship between news and market activity (MV Volume and Absolute Order Imbalance) is particularly strong. Interestingly, there is also a significant positive relationship between the news dummy 60 seconds prior to the actual news arrival and variables for market activity for MV Volume, Absolute Order Imbalance, Volatility and Spreads. This is suggestive of market participants becoming aware of the imminent arrival of non-scheduled news and reacting in advance. Once the news item has been released, the direct impact of news as captured by the dummy variables disappears quickly.

Table 3 reports the results of the Variance Error Decomposition[9] relating to the VAR specification in Eq. (3); once again average results for the 33 stocks in the sample are reported. In general, the own-variable lags contribute most to the variance error for each variable (in the range of 89-100%); for Volume and Returns this is close to 100%. It is noticeable that Volume contributes to the variance error of volatility (2.5%), order imbalance (8.5%), and average trade size (50%). Volatility is the only other variable to influence the variance error of others in this VAR specification (1.8% for Order Imbalance and 8% for Bid-Ask Spread) indicating the importance of both Volume and Volatility in driving the other standardised microstructure variables considered here.

<Insert Table 3>

Impulse response analysis is performed in order to provide greater insights into news-implied market responses. A news impulse is defined as a change in the corresponding news dummy. Fig. 9 depicts the impulse response to news-induced changes based on the averaged VAR estimates. The general finding is that the response to news impulse subsides quickly and is

[9] Belsley, Kuh and Welsch (2004) method is used.

insignificant after five 30 sec intervals. The response of returns to news has the greatest magnitude, but also subsides most quickly.

<Insert Figure 9>

Overall, the dynamic analysis confirms the unconditional effects in Section 3. Although all variables exhibit a significant reaction to news arrival, the most sensitive measures are those concerning market activity and volatility.

5. Conclusion

Analysing the impact of news on a specific financial asset is challenging since the amount of news and the speed of information dissemination has increased rapidly over the past 20 years. The prodigious amount of information creates a significant amount of noise alongside the actual release of news. Such effects have made it difficult to identify a genuine linkage between trading activity and the intraday news flow. Previous studies have tended to focus on scheduled and homogenous types of news, such as earnings and macroeconomic announcements, with non-scheduled news being virtually ignored. In contrast, this paper has sought to explore the relationship between non-scheduled news-flow and stock market activity in an Australian context, over a lengthy 12-year sample period.

Unconditional analysis reveals that high-relevance news items induce an increase in market activity, volatility and spreads. Disaggregating news items by sentiment demonstrates that negative news has a greater impact than positive news, whilst neutral news induces no impact. There is evidence that prices react to news prior to its arrival on the Dow Jones news-wire but the level of returns is not sufficient to be profitable after considering transaction costs.

An investigation into the behaviour of the Big 4 banking stocks (ANZ, CBA, NAB, Westpac) produces similar results, and also reveals that the impact of news on market activity and volatility is greater following the onset of the GFC. Indeed, prior to the GFC the relationship between news and returns completely disappears.

Analysis of market dynamics and cross-dependencies between variables in a high-frequency VAR model confirms the significant market impact induced by contemporaneous news items; with a particularly strong relationship between news and market activity measures. This framework also reveals a significant and positive relationship between the measures of trading activity and volatility. Finally, there is an increase in bid-ask spreads if prior trading periods reflect rising market activity and volatility. The results are consistent with existing microstructure theory, corroborate the empirical evidence presented by Groß-Klußmann and Hautsch (2011) for the London stock market, and are thus suggestive of the results been applicable in a global context.

The results have implications for traders in ASX stocks, and equity markets in general. Specifically, the results suggest that linguistic processing software, such as that provided by Ravenpack, is able to successfully categorise high-frequency news; this may allow traders, particularly high-frequency traders, to benefit from the implementation of algorithmic trading strategies involving such software. In addition, the speed with which markets react to news may also hint at policy implications in terms of ensuring the timely release of news stories to all market participants.

Further analysis could extend this work to other countries to determine whether the findings are Australian specific or are general results which are applicable in an international context. In addition, investigating the impact of news on less liquid stocks would be of interest as ex-ante one would expect fewer news releases concerning such stocks and a resulting increase in the impact of news of any kind. It may also be possible to determine from ASX data whether activity initiated by different types of brokers (e.g. retail and institutional) results in different levels of return. Finally, one may wish to explore the information linkages between news, cash-stock markets and corresponding derivative markets.

Acknowledgment

Transaction level ASX stock data provided by SIRCA.

The author would like to thank participants at the 25th Australasian Finance & Banking Conference (2012) in Sydney for useful comments in preparing this article.

References

Antweiler, W., and M.Z. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance*, 59, 1259-1294

Bai,J., and P. Perron, 2003, Computation and analysis of multiple structural change models, *Journal of Applied Econometrics*, 18, 1-22

Becker, K., J. Finnerty, and K. Kopecky, 1996, Macroeconomic news and the efficiency of international bond futures markets, *Journal of Futures Markets*, 16, 131-145

Bernanke, B., and K. Kuttner, 2005, What explains the stock market's reaction to Federal Reserve policy?, *Journal of Finance*, 60, 1221-1257

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, 1997, The Econometrics of Financial Markets, Princeton University Press

Chen, C.W.S., T.C. Chiang, and M.K.P. So, 2003, Asymmetrical reaction to US stock-return news: evidence from major stock markets based on a double-threshold model, *Journal of Economics and Business*, 55, 487-502

De Bondt, W.F.M., and R. Thaler, 1985, Does the stock market overreact?, *Journal of Finance*, 40, 793-805

Dzielinski, M., 2011, News sensitivity and the cross-section of stock returns, NCCR Finrisk working paper no. 719

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

Easley, D., and M. O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance*, 47, 577-605

Ederington, L., and J. Lee, 1993, How markets process information: news releases and volatility, *Journal of Finance*, 48, 1161-1191

Ederington, L., and J. Lee, 1995, The short-run dynamics of the price adjustment to new information, *Journal of Financial and Quantitative Analysis*, 30, 117-134

Engelberg, J.E., A.V. Reed, and M.C. Ringgenberg, 2012, How are shorts informed? Short sellers, news and information processing, *Journal of Financial Economics*, Available online March 2012

Groß-Klußmann, A., and N. Hautsch, 2011, When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions, *Journal of Empirical Finance*, 18, 321-340

Harris, M., and A. Raviv, 1993, Differences of opinion make a horse race, *Review of Financial Studies*, 6, 473-506

Kyle, A., 1985, Continuous auctions and insider trading, *Econometrica*, 53, 1315-1335

Lee, C., and M. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance*, 46, 733-746

Patell, J., and M. Wolfson, 1984, The intraday speed of adjustment of stock prices to earnings and dividend announcements, *Journal of Financial Economics*, 13, 223-252

Rigobon, R., and B. Sack, 2004, The impact of monetary policy on asset prices, *Journal of Monetary Economics*, 51, 1553-1575

Sinha, N.R., 2011, Under-reaction to news in the US stock market, Working Paper

Smales, L.A., 2013, Impact of macroeconomic announcements on interest rate futures: High-frequency evidence from Australia, *Journal of Financial Research*, 36, 371-388

Tetlock, P.C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance*, 62, 1139-1168

Tetlock, P.C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance*, 63, 1437-1467

Tetlock, P.C., 2011, All the news that's fit to reprint: Do investors react to stale information, *Review of Financial Studies*, 24, 1481-1512

Woodruff, C.S., and A.J. Senchack, 1988, Intraday price-volume adjustments of NYSE stocks to unexpected earnings, *Journal of Finance*, 43, 467-491

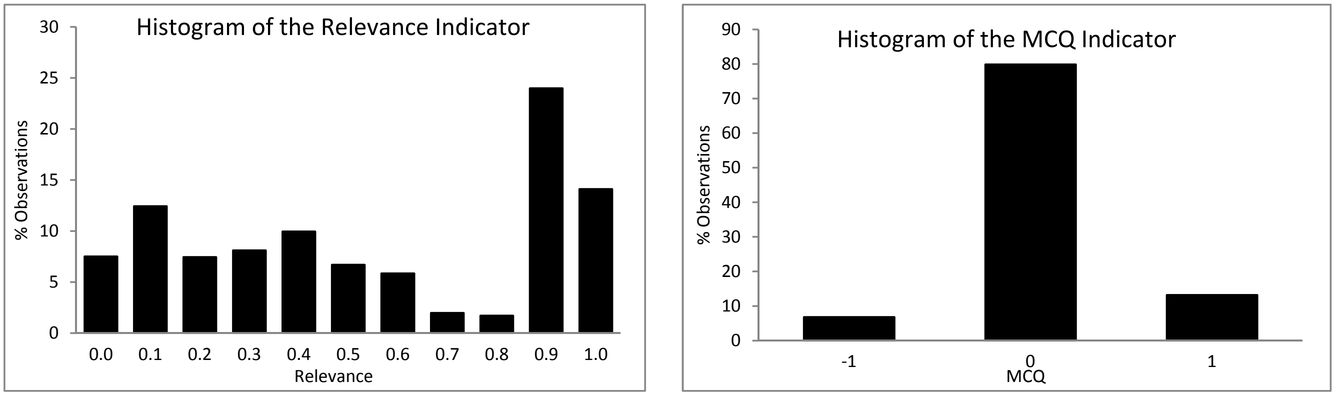

Fig. 1a. Distribution of relevance indicator and MCQ indicator. Relevance > 0.9 indicates highly relevant news. MCQ of -1, 0 & +1 indicates highly relevant news that is determined to be negative, netural positive respectively.

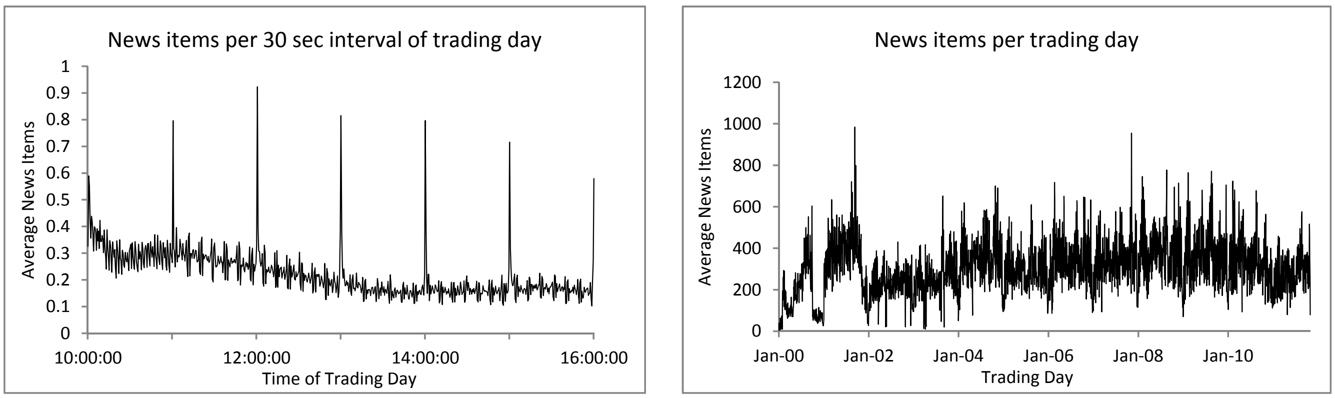

Fig. 1b. Distribution of news items over the 30 second intervals of a trading day, and distribution of news items per trading day.
Period: 04 Jan 2000 - 01 Nov 2011

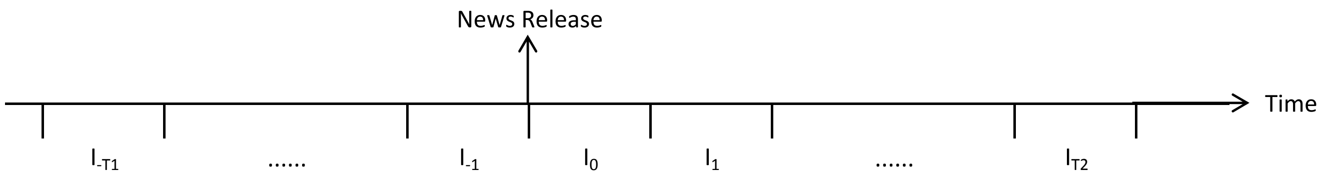

Fig. 2. Intervals around news arrival

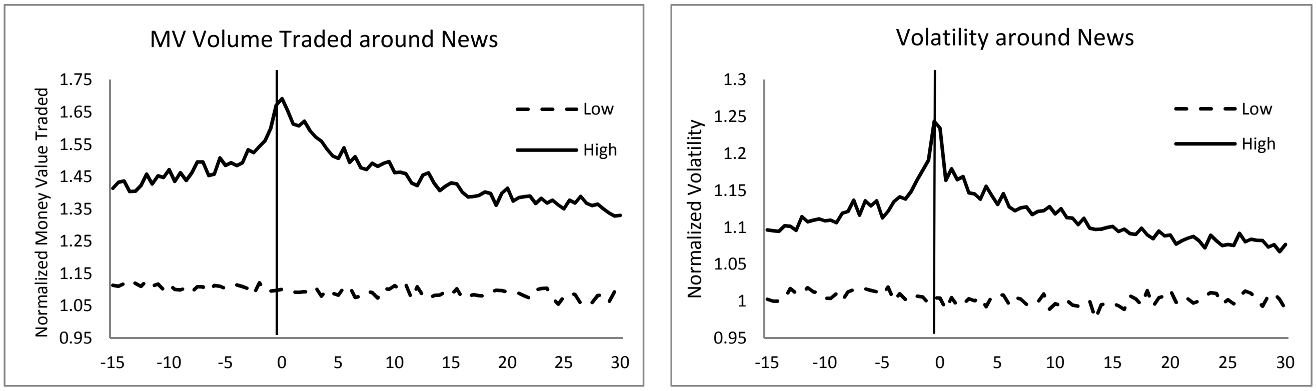

Fig. 3a. Money Value Traded and Volatility around low and high relevance news releases.

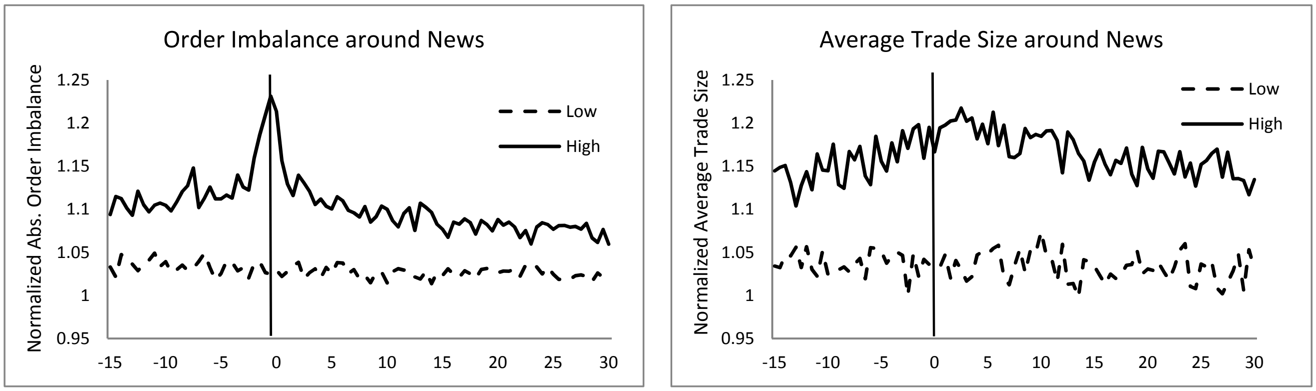

Fig. 3b. Order Imbalance and Average Trade Size around low and high relevance news releases.

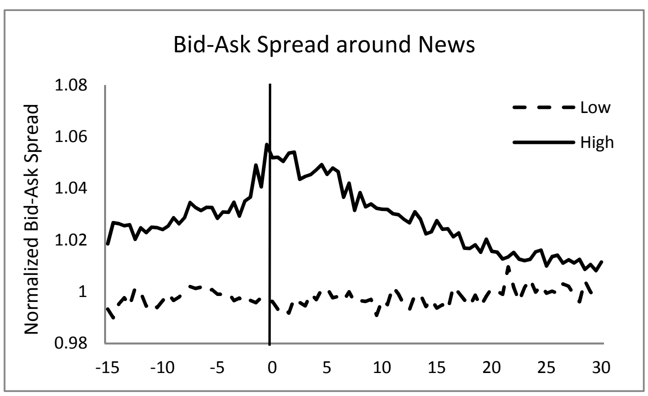

Fig. 3c. Bid-Ask Spread around low and high relevance news releases.

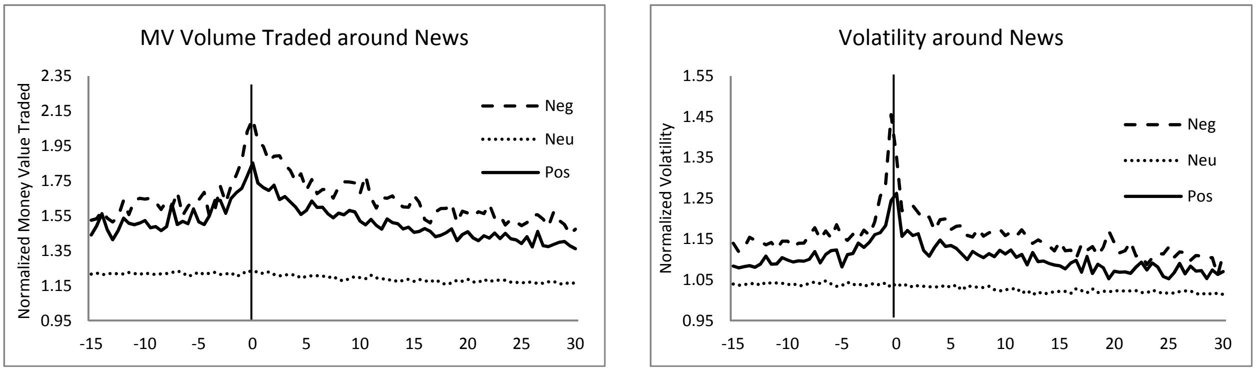

Fig. 4a. Money Value Traded and Volatility around positive and negative news releases.

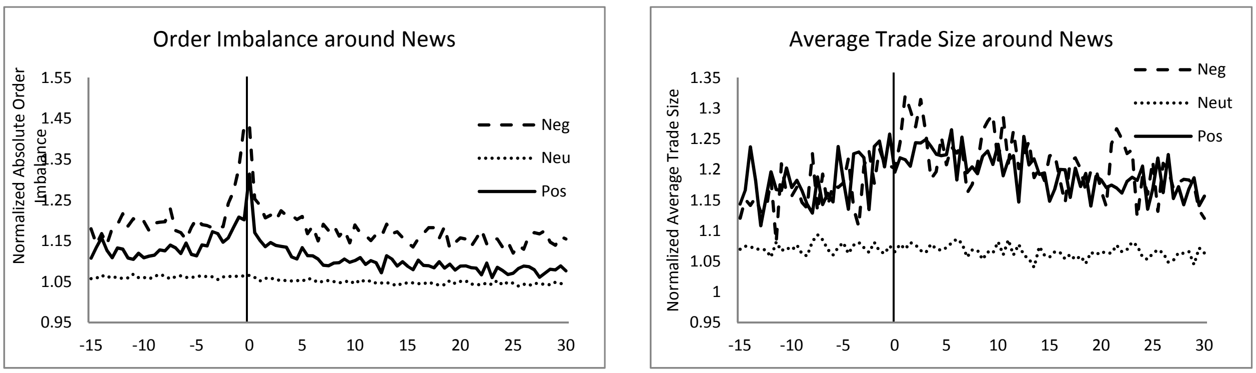

Fig. 4b. Order Imbalance and Average Trade Size around positive and negative news releases.

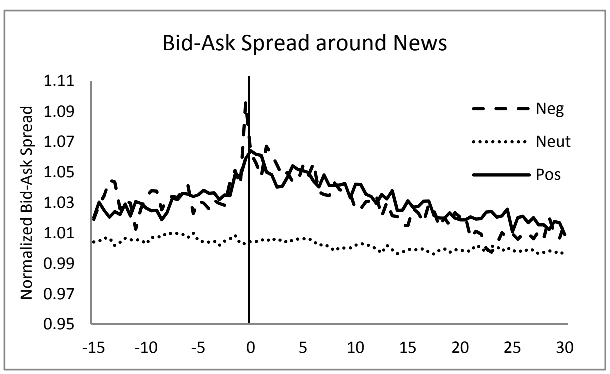

Fig. 4c. Bid-Ask Spread around positive and negative news releases.

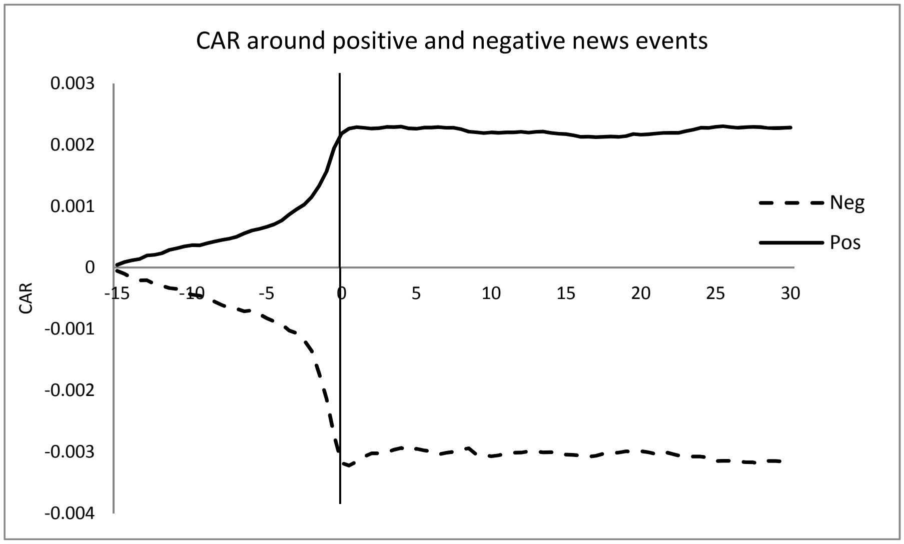

Fig. 5. Cumulated Abnormal Returns around highly relevant positive and negative news.

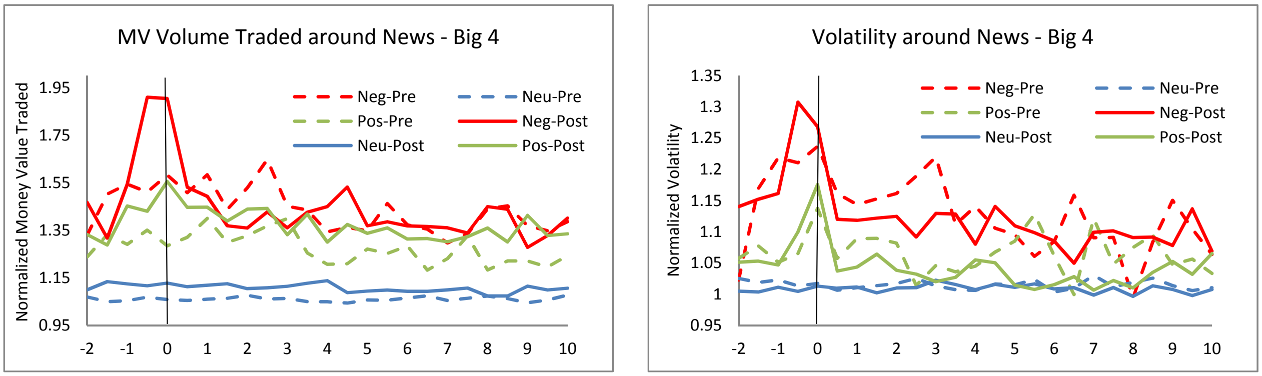

Fig. 6a. Money Value Traded and Volatility around positive and negative news releases, disaggregated by pre- and post-GFC periods.

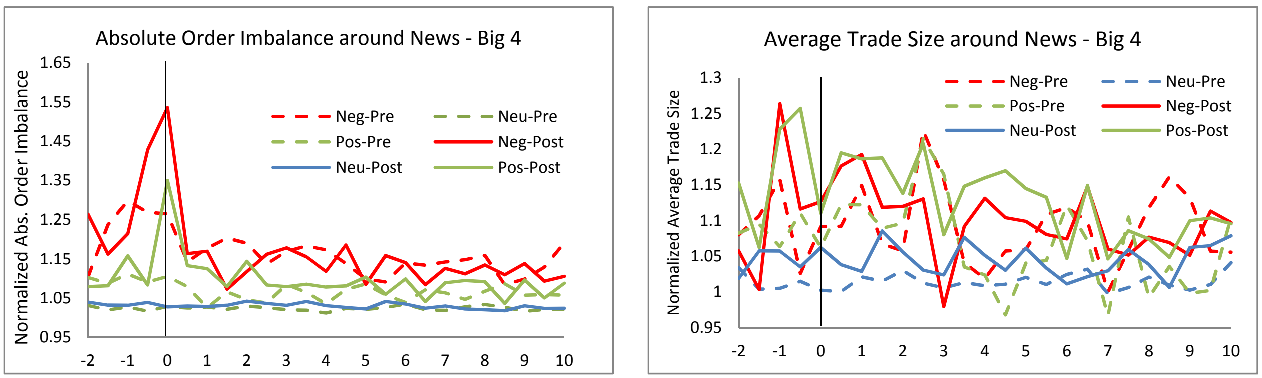

Fig. 6b. Order Imbalance and Average Trade Size around positive and negative news releases, disaggregated by pre- and post-GFC periods.

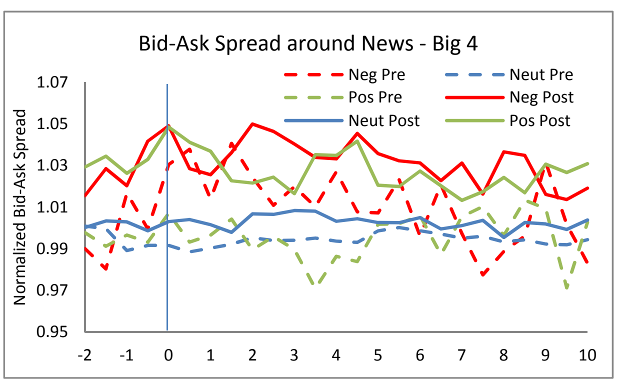

Fig. 6c. Bid-Ask Spread around positive and negative news releases, disaggregated by pre- and post-GFC periods.

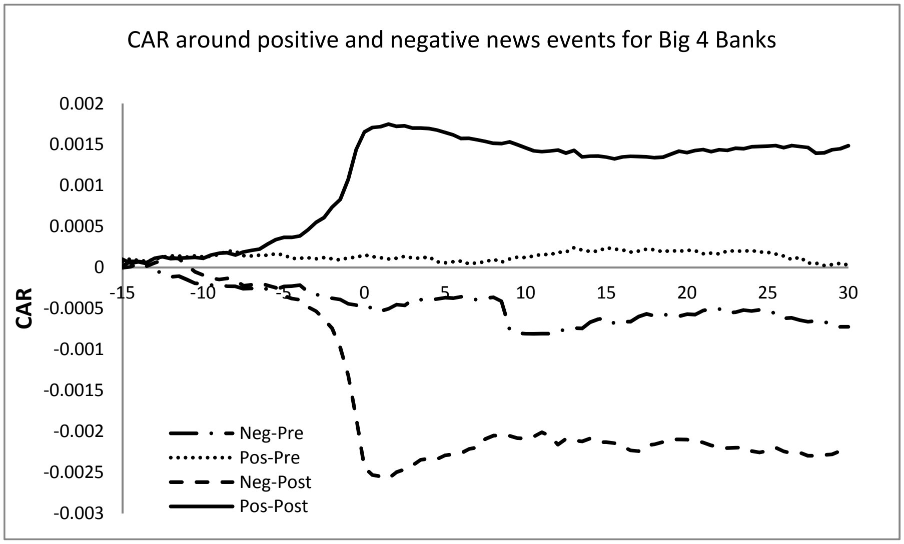

Fig. 7. Cumulated Abnormal Returns around highly relevant positive and negative news, disaggregated into pre-GFC and post-GFC periods for the Big 4 Australian Banks.

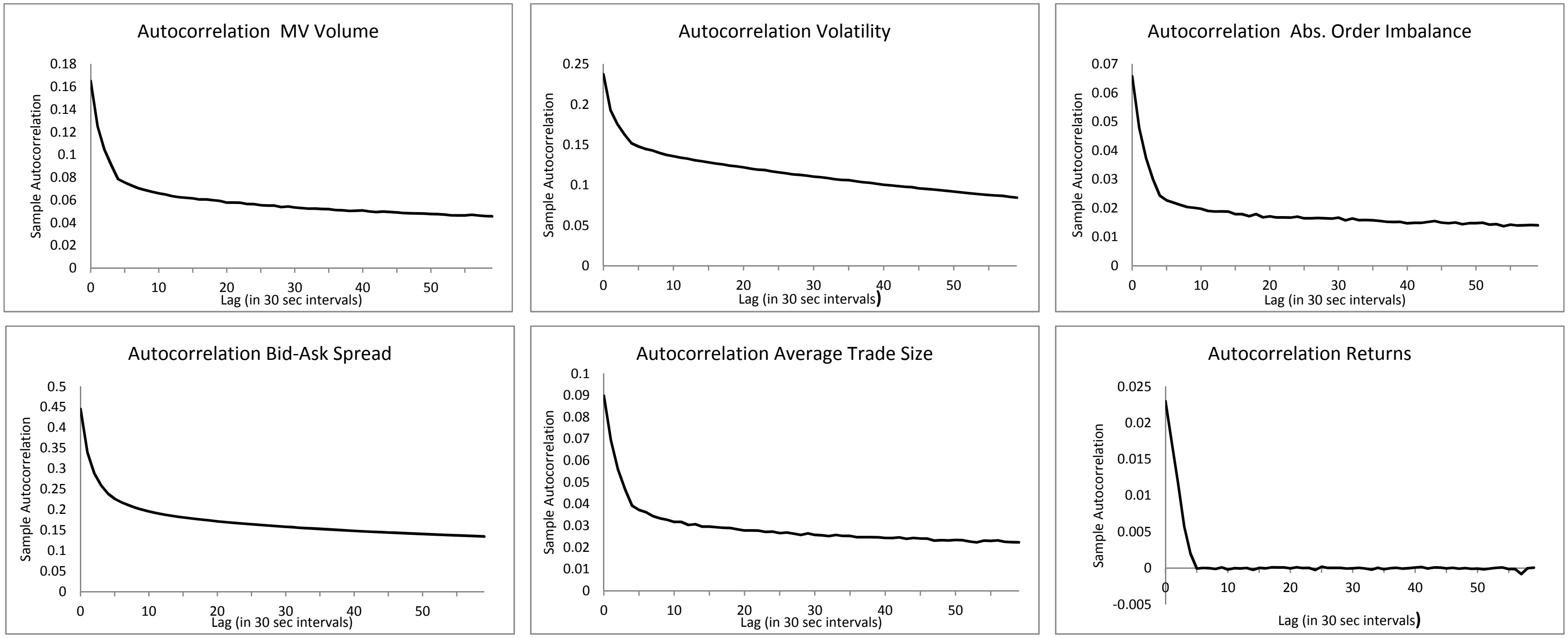

Fig 8. Autocorrelation functions for the normalized variables of interest (Volume, Volatility, Abs. Order Imbalance, Bid-Ask Spread, Ave. Trade Size, Returns)

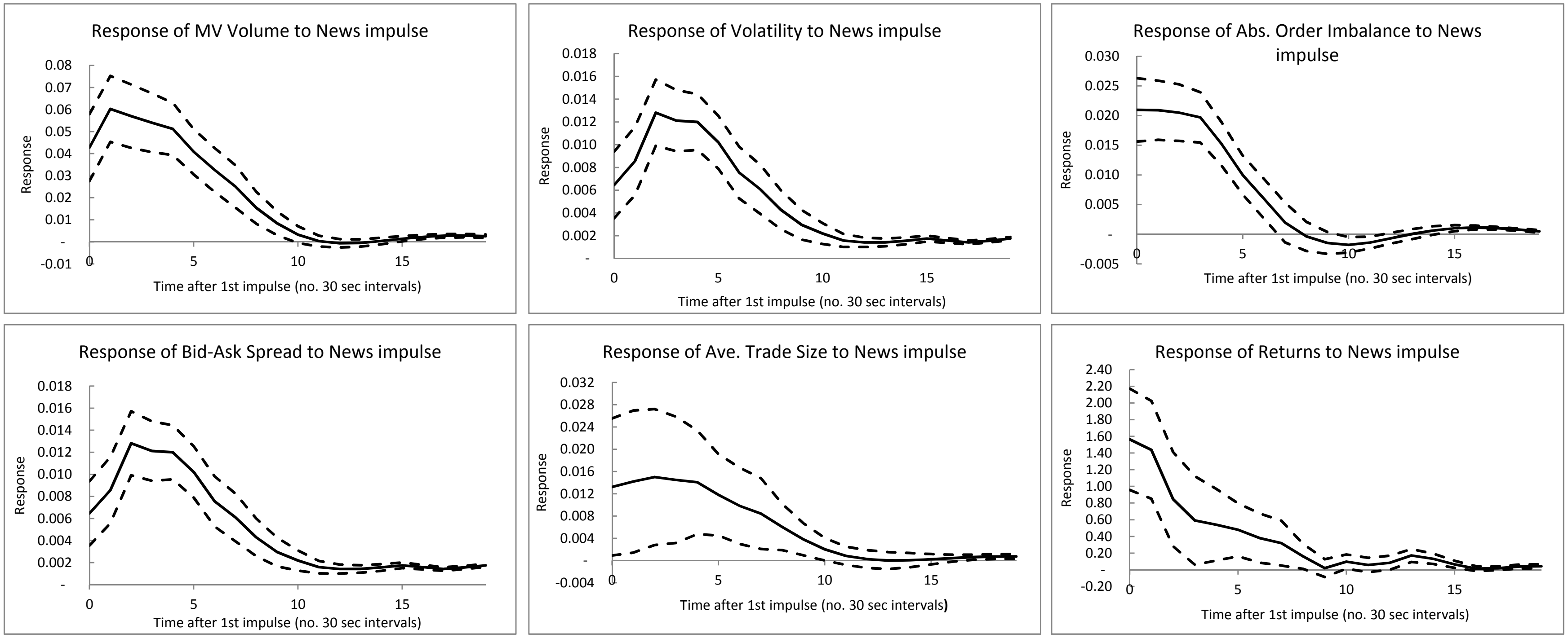

Fig 9. Response analysis of a change in the news dummy for highly relevant news items

Table 1
Descriptive Statistics

RIC	GICS	M. Cap in % of All Ord Index	Money Value Traded	Return	Spread	No. Trades	No. News Items	No. High Relevance News Items	Neg. MCQ	Neut. MCQ	Pos. MCQ
AMC	15	0.75%	22,472,726	-0.002	0.016	1,265	3,464	1,856	274	1,204	378
AMP	40	1.00%	53,331,876	-0.042	0.015	1,985	4,221	2,534	338	1,596	600
ANZ	40	5.01%	129,650,240	0.030	0.018	3,878	28,499	9,372	678	7,698	996
ASX	40	0.47%	16,230,932	0.044	0.039	1,442	12,496	2,224	172	1,774	278
BHP	15	15.20%	366,157,877	0.029	0.019	6,189	51,401	16,478	1,136	13,138	2,204
CBA	40	6.58%	149,347,273	0.025	0.026	3,986	37,172	7,050	272	6,337	441
CCL	30	0.79%	19,967,333	0.038	0.018	1,223	4,793	1,953	137	1,470	346
CPU	45	0.39%	13,120,398	0.013	0.019	1,215	14,169	1,410	68	1,267	75
CSL	35	1.56%	46,088,345	0.020	0.040	2,333	2,075	1,435	226	633	576
GPT	40	0.48%	26,559,975	0.007	0.013	828	1,783	1,253	223	771	259
ILU	15	0.59%	6,336,857	0.032	0.020	728	2,759	1,103	250	503	350
LEI	20	0.59%	15,789,150	0.063	0.042	1,562	3,910	2,610	255	1,321	1,034
LLC	40	0.31%	20,353,338	-0.034	0.034	1,130	4,453	2,330	255	1,616	459
MGR	40	0.35%	16,671,761	-0.035	0.014	763	1,365	887	181	487	219
NAB	40	4.60%	165,679,643	0.002	0.021	4,017	53,669	16,511	738	14,830	943
NCM	15	1.82%	46,430,777	0.078	0.031	2,406	17,610	4,637	191	4,100	346
ORG	10	1.20%	22,597,322	0.099	0.020	1,556	4,318	2,051	166	1,561	324
ORI	15	0.82%	21,450,359	0.042	0.032	1,527	3,377	1,568	176	1,077	315
OSH	10	0.77%	18,467,148	0.005	0.015	1,037	773	572	99	271	202
QAN	20	0.31%	45,211,832	-0.015	0.011	1,283	25,590	9,771	751	7,962	1,058
QBE	40	1.26%	56,040,752	0.034	0.025	2,485	2,887	1,582	200	982	400
RIO	15	10.13%	151,216,208	0.036	0.044	3,769	50,614	13,534	651	11,653	1,230
SGP	40	0.58%	26,166,113	0.005	0.018	1,039	2,633	1,113	104	826	183
SHL	35	0.40%	10,731,611	0.019	0.026	1,084	1,103	722	96	352	274
STO	10	1.10%	32,335,869	0.043	0.022	1,864	3,860	2,805	283	1,489	1,033
SUN	40	0.88%	33,954,597	-0.023	0.021	1,739	3,137	1,091	92	847	152
TCL	20	0.68%	16,216,459	0.019	0.017	934	1,295	1,006	195	482	329
TLS	50	3.50%	142,819,515	-0.040	0.010	2,546	39,706	15,659	1,561	13,157	941
TOL	20	0.34%	19,242,407	0.002	0.024	1,235	3,116	1,767	148	1,096	523
WBC	40	5.54%	126,628,123	0.029	0.018	3,685	72,087	20,639	664	19,036	939
WES	20	2.82%	52,597,824	0.036	0.028	2,370	7,446	3,044	269	2,392	383
WOW	30	2.62%	60,957,963	0.062	0.022	2,555	7,680	2,659	163	2,131	365
WPL	10	2.29%	66,161,104	0.050	0.036	2,718	10,979	4,452	540	3,223	689
Sum		75.72%					484,440	157,678	11,552	127,282	18,844

RIC denotes the Reuters Identification Code for each stock. GICS is the Global Industry Classification Standard. % M. Cap of All Ord Index is the % market capitalisation, defined as the share price multiplied by the number of ordinary shares in issue for 2011. Money value traded is computed as the trade size times the respective price, figure reported is the average daily value over the sample period (in 1,000). Return is the average daily return, defined as log(Pt/Pt-1), x100 over the sample period. Spread and No. Trades are averages per trading day. The no. news items is the total number of news items per firm without overnight news and entries where the event index > 2. High relevance news items are the total number of news items with a Relevance indicator > 90. Neg. MCQ, Neut. MCQ, Pos. MCQ provide the number of relevant negative, neutral, and positive news items respectively.

Sample Period: Jan '00 - Nov '11

Table 2
Average VAR results

		MV Volume		Volatility		Abs. Order Imbalance		Bid Ask Spread		Ave. Trade Size		Returns	
Dynamics													
	c	0.788	***	0.330	***	0.719	***	0.382	***	0.799	***	3.532	
		(0.024)		(0.012)		(0.010)		(0.005)		(0.023)		(21.588)	
MV Volume	$Volume_{t-1}$	0.225	***	0.022	***	0.023	***	0.005	***	0.066	***	0.189	
		(0.004)		(0.002)		(0.002)		(0.001)		(0.004)		(3.967)	
	$Volume_{t-2}$	0.134	***	0.012	***	0.012	***	-0.001		0.039	***	0.501	
		(0.004)		(0.002)		(0.002)		(0.001)		(0.004)		(3.940)	
	$Volume_{t-3}$	0.099	***	-0.002		0.008	***	-0.002	*	0.029	***	-0.975	
		(0.004)		(0.002)		(0.002)		(0.001)		(0.004)		(3.898)	
Volatility	Vol_{t-1}	0.014	**	0.218	***	-0.004		0.037	***	0.028	***	-1.026	
		(0.005)		(0.003)		(0.003)		(0.001)		(0.006)		(5.260)	
	Vol_{t-2}	0.011	*	0.128	***	0.006	**	0.009	***	0.005		-0.533	
		(0.006)		(0.003)		(0.003)		(0.001)		(0.006)		(5.291)	
	Vol_{t-3}	0.006		0.109	***	0.004		0.005	***	0.001		0.035	
		(0.006)		(0.003)		(0.003)		(0.001)		(0.006)		(5.263)	
Abs. Order Imbalance	OI_{t-1}	0.037	***	0.030	***	0.130	***	0.011	***	0.020	***	2.090	
		(0.007)		(0.004)		(0.003)		(0.002)		(0.007)		(6.749)	
	OI_{t-2}	0.013	*	0.004		0.078	***	0.000		-0.013	*	3.612	
		(0.007)		(0.004)		(0.003)		(0.002)		(0.007)		(6.661)	
	OI_{t-3}	0.001		-0.003		0.054	***	-0.001		-0.012	*	-0.935	
		(0.007)		(0.004)		(0.003)		(0.001)		(0.007)		(6.488)	
Bid-Ask Spread	BAS_{t-1}	-0.294	***	0.064	***	-0.206	***	0.365	***	-0.161	***	4.311	
		(0.015)		(0.008)		(0.006)		(0.003)		(0.014)		(13.341)	
	BAS_{t-2}	0.035	**	0.024	***	-0.026	***	0.098	***	0.021	**	-3.302	
		(0.016)		(0.008)		(0.007)		(0.003)		(0.010)		(13.971)	
	BAS_{t-3}	0.013		-0.004		0.011	*	0.050	***	-0.005	***	5.059	
		(0.016)		(0.008)		(0.007)		(0.003)		(0.015)		(13.888)	
Ave. Trade Size	$AvTrade_{t-1}$	-0.088	***	-0.015	***	-0.018	***	-0.003	***	0.029	***	0.038	
		(0.004)		(0.002)		(0.002)		(0.001)		(0.004)		(3.791)	
	$AvTrade_{t-2}$	-0.039	***	-0.008	***	-0.010	***	0.000		0.035	***	-0.728	
		(0.004)		(0.002)		(0.002)		(0.001)		(0.004)		(3.821)	
	$AvTrade_{t-3}$	-0.026	***	-0.001		-0.006	***	0.001		0.028	***	1.382	
		(0.004)		(0.002)		(0.002)		(0.001)		(0.004)		(3.865)	
Period Returns	Ret_{t-1}	-0.0002		-0.0001		0.0000		0.0000		0.0001		0.033	***
		(0.0009)		(0.0004)		(0.0004)		(0.0002)		(0.0009)		(0.003)	
	Ret_{t-2}	0.0007		0.0002		0.0000		-0.0001		0.0006		0.022	***
		(0.0009)		(0.0004)		(0.0004)		(0.0002)		(0.0009)		(0.003)	
	Ret_{t-3}	-0.0003		0.0002		0.0001		0.0000		-0.0003		0.017	***
		(0.0009)		(0.0004)		(0.0004)		(0.0002)		(0.0009)		(0.003)	
News Dummies													
Dummy leads	$News_{t+2}$	0.283	**	0.061	***	0.113	**	-0.015		0.059		-8.132	
		(0.119)		(0.065)		(0.047)		(0.024)		(0.119)		(118.18)	
	$News_{t+1}$	0.274	**	0.062		0.138	**	0.025	**	0.094		-15.111	
		(0.132)		(0.078)		(0.056)		(0.011)		(0.143)		(142.10)	
Item dummy	$News_{t}$	0.313	**	0.095	**	0.165	***	0.028	***	0.108	***	-9.410	
		(0.135)		(0.044)		(0.053)		(0.008)		(0.036)		(136.85)	
Dummy lags	$News_{t-1}$	0.228	**	0.069	***	0.122	**	0.052	**	0.126	***	1.098	
		(0.112)		(0.020)		(0.050)		(0.026)		(0.045)		(129.49)	
	$News_{t-2}$	0.086		0.000		0.029		-0.019		0.018		7.328	
		(0.133)		(0.073)		(0.053)		(0.028)		(0.134)		(135.57)	
	$News_{t-3}$	0.069		0.021		0.035		0.008		0.014		-6.437	
		(0.137)		(0.075)		(0.054)		(0.028)		(0.138)		(137.79)	
	$News_{t-4}$	-0.015		0.000		-0.017		0.008		0.095		10.877	
		(0.125)		(0.062)		(0.044)		(0.023)		(0.113)		(112.73)	

The table provides estimation results for the VAR model outlined in equation 3. Estimates are provided for the dynamics of the endogenous variables, together with the exogenous news dummies. Reported coefficients are averages of the estimates for each individual stock with standard errors given in the parantheses below. Significance is reported based on average t-statistics. *** denotes significance of the average coefficient estimates at the 1% level, ** at the 5% level, and * at the 10% level.

Table 3

Average Variance Error Decomposition results

MV Volume

Lag Period	Standard Error	MV Volume	Volatility	Abs. Order Imbalance	Bid-Ask Spread	Ave. Trade Size	Period Returns
1	2.381	100.000	0.000	0.000	0.000	0.000	0.000
2	2.433	99.103	0.008	0.098	0.398	0.389	0.004
3	2.466	98.769	0.013	0.150	0.467	0.587	0.014
4	2.494	98.514	0.019	0.184	0.512	0.756	0.014
5	2.518	98.304	0.023	0.206	0.535	0.917	0.015
6	2.528	98.172	0.025	0.224	0.566	0.998	0.015

Volatility

Lag Period	Standard Error	MV Volume	Volatility	Abs. Order Imbalance	Bid-Ask Spread	Ave. Trade Size	Period Returns
1	1.273	2.224	97.776	0.000	0.000	0.000	0.000
2	1.309	2.452	97.300	0.082	0.123	0.040	0.004
3	1.332	2.515	97.146	0.108	0.170	0.054	0.006
4	1.354	2.532	97.091	0.116	0.189	0.063	0.010
5	1.379	2.502	97.053	0.115	0.251	0.066	0.013
6	1.389	2.510	96.988	0.119	0.300	0.070	0.013

Abs. Order Imbalance

Lag Period	Standard Error	MV Volume	Volatility	Abs. Order Imbalance	Bid-Ask Spread	Ave. Trade Size	Period Returns
1	1.028	8.408	1.786	89.806	0.000	0.000	0.000
2	1.049	8.489	1.776	88.553	1.092	0.088	0.002
3	1.059	8.595	1.767	88.252	1.241	0.142	0.002
4	1.065	8.689	1.761	88.041	1.319	0.186	0.003
5	1.069	8.773	1.758	87.887	1.353	0.225	0.004
6	1.070	8.809	1.760	87.778	1.403	0.246	0.004

Bid-Ask Spread

Lag Period	Standard Error	MV Volume	Volatility	Abs. Order Imbalance	Bid-Ask Spread	Ave. Trade Size	Period Returns
1	0.523	0.140	5.856	0.179	93.825	0.000	0.000
2	0.567	0.301	7.835	0.264	91.581	0.012	0.007
3	0.586	0.364	8.817	0.291	90.504	0.016	0.008
4	0.596	0.391	9.540	0.308	89.729	0.018	0.014
5	0.605	0.398	10.196	0.304	89.065	0.019	0.017
6	0.611	0.410	10.726	0.306	88.520	0.020	0.017

Ave. Trade Size

Lag Period	Standard Error	MV Volume	Volatility	Abs. Order Imbalance	Bid-Ask Spread	Ave. Trade Size	Period Returns
1	2.419	49.519	0.274	3.132	0.023	47.053	0.000
2	2.433	49.743	0.277	3.131	0.144	46.703	0.002
3	2.443	49.964	0.277	3.127	0.159	46.465	0.009
4	2.451	50.146	0.280	3.126	0.173	46.264	0.011
5	2.457	50.319	0.281	3.124	0.183	46.081	0.011
6	2.459	50.364	0.282	3.122	0.193	46.029	0.011

Period Returns

Lag Period	Standard Error	MV Volume	Volatility	Abs. Order Imbalance	Bid-Ask Spread	Ave. Trade Size	Period Returns
1	1412.19	0.001	0.014	0.003	0.003	0.001	99.979
2	1417.33	0.002	0.015	0.004	0.005	0.001	99.974
3	1421.30	0.003	0.016	0.004	0.006	0.002	99.969
4	1423.74	0.004	0.017	0.005	0.007	0.004	99.963
5	1425.00	0.004	0.019	0.006	0.007	0.005	99.959
6	1425.44	0.005	0.019	0.006	0.007	0.005	99.958

The table provides results for the variance error decomposition relating to the VAR model specified in equation 3 (the results of this model are reported in Table 2).