COVER SHEET

Predicting the Short-Term Market Reaction to Asset Specific News: Is Time Against Us?

Calum Robertson[1], Shlomo Geva[1], Rodney Wolff[2]

{[1] Information Research Group, Queensland University of Technology
[2] School of Economics and Finance, Queensland University of Technology}
2 George Street, Brisbane, Queensland Australia 4000
{cs.robertson, s.geva, r.wolff}@qut.edu.au

Abstract. The efficient market hypothesis states that investors immediately incorporate all available information into the price of an asset to accurately reflect its value at any given time. The sheer volume of information immediately available electronically makes it difficult for a single investor to keep abreast of all information for a single stock, let alone multiple. We aim to determine how quickly investors tend to react to asset specific news by analysing the accuracy of classifiers which take the content of news to predict the short-term market reaction. The faster the market reacts to news the more cost-effective it becomes to employ content analysis techniques to aid the decisions of traders. We find that the best results are achieved by allowing investors in the US 90 minutes to react to news. In the UK and Australia the best results are achieved by allowing investors 5 minutes to react to news.

Keywords. Document Classification, Stock Market, News, SVM, C4.5.

1. Introduction

Not so long ago most traders relied upon newspapers and magazines to supply the information they required to invest. However the last two decades has seen a rapid increase in the availability of both real-time prices and media coverage via electronic sources. It has become difficult for a single person to keep abreast of all available information for a single asset, and impossible for multiple [1].

There is a plethora of research which shows that specific asset markets react to news. This includes the reaction to newspaper, magazine, and other real-time sources such as websites [2-7]. There is also plenty of evidence that markets react to macroeconomic announcements (information released by governments to provide an indication of the state of the local economy) [8-14]. It has also been shown that investors take heed of analyst recommendations [15-17] to aid their investing decisions. Interestingly it has also been shown that the futures prices of oranges are influenced by the release of weather reports [18].

Ederington and Lee found that volatility on Foreign Exchange and Interest Rate Futures markets increases within one minute of a macroeconomic news announcement, and the effect lasts for about 15 minutes [8]. They later determined

that the same markets begin to react to news within 10 seconds of macroeconomic news announcements, with weak evidence that they tend to overreact to news within the first 40 seconds after news, but settle within 3 minutes [9]. Graham et al. established that the value of stocks on the S&P 500 index is influenced by scheduled macroeconomic news, however, they didn't investigate any intraday effect [12]. Nofsinger and Prucyk concluded that unexpected bad macroeconomic news is responsible for most abnormal intraday volume trading on the S&P 100 Index option [14].

Despite strong evidence that the stock market does react to macroeconomic news, there is far more asset specific news than macroeconomic news. Furthermore, unlike macroeconomic news, most asset specific news isn't scheduled and therefore investors have not formed their own expectation, or adopted analysts' recommendations about the content of the news.

Wutherich et al. analysed the content of news available on a well known website and used it to predict what the given index would do on the next day [19]. They achieved a statistically significant level of accuracy on their forecasts, though it is somewhat easier to predict the direction of an index in a day than it is to predict the reaction of a single asset in less than a day.

Fung et al. examined the effect of all asset specific news on a limited number of stocks and found that they could make money based on the predictions of a system which processed the content of the news [20]. However they didn't report the classification accuracy of their system so it is difficult to determine how good their results are.

Mittermayer investigated the effect of Press Announcements on the New York Stock Exchange and the NASDAQ and determined that the content of news can be used to predict, with reasonable accuracy, when the market will experience high returns within an hour of the announcement [7]. Unfortunately press announcements are only a fraction of asset specific news, so further investigation is required to determine how the stock market reacts, if at all, to this type of news.

It is vitally important to examine how rapidly the market reacts to asset specific news, in order to capitalise on its content. The faster the market reacts to news the more cost-effective it becomes to employ content analysis techniques to aid the decisions of traders. Conversely if the market is slow to react to asset specific news then there is little point in utilising content analysis to highlight interesting articles.

In this paper we formulate a methodology to determine how quickly the stock market reacts to news. We apply this to stocks traded in the US, UK and Australian markets to ascertain how strongly these markets react, if at all, to asset specific news. Firstly we describe the data used in our tests, and the methodology we employed, before presenting our results and conclusions.

2. Data

All data for this research was obtained using the Bloomberg Professional® service. The dataset consists of stocks which were in the S&P 100, FTSE 100, and ASX 100 indices as at the 1st of July 2005 and continued to trade through to the 1st of November

2006, which is a total of 283 stocks. For each stock the Trading Data, and News were collected for the period beginning 1st of May 2005 through to and including the 31st of October 2006.

The set defined in Eq. (1) consists of each distinct minute where trading occurred for the stock (s), within all minutes for the period of data collection (T_A), and the average price for trades during that minute. However we are only interested in the business time scale (minutes which occurred during business hours for the market on which the stock trades). Furthermore we want a homogeneous time series (i.e. an entry for every business trading minute for the stock, regardless of whether any trading occurred). Therefore we produce the date and price time series for all minutes in the business time scale (T_B) with the definitions in Eqs. (2)-(3), where we define the price at time t as the price of the last actual trade for the stock prior to or at the given time. Note that if the stock was suspended from trading for a whole day then the day is excluded from T_B.

$$I_{(s)} = \{I_1, I_2, \ldots, I_m\} \mid I_{(s,z)} = (d_{(s,z)}, p_{(s,z)}) \wedge z \in T_A \tag{1}$$

$$D_{(s)} = \{D_1, D_2, \ldots, D_n\} \mid D_{(s,t)} > D_{(s,t-1)} \wedge D_{(s,t)} \in T_B \wedge T_B \subseteq T_A \tag{2}$$

$$P_{(s)} = \{P_1, P_2, \ldots, P_n\} \mid P_{(s,t)} = (p_{(s,z)} \mid z = \max(z \mid d_{(s,z)} \leq D_{(s,t)})) \tag{3}$$

The news search facility within the Bloomberg Professional® service was used to download all relevant articles for each stock within the dataset. These articles include Press Announcements, Annual Reports, Analyst Recommendations and general news which Bloomberg has sourced from over 200 different news providers.

The set defined in Eq. (4) consists of each distinct news article for the stock, which occurred during business hours excluding the first and last Δt minutes of the day, and contains the time and content of the article. The first Δt minutes are excluded as the market behaves differently during this period and therefore this could skew the results of a classifier. The last Δt minutes are excluded as the market doesn't have time to react to the news within the period, and this could also skew the results of a classifier.

$$A_{(s)} = \{A_1, A_2, \ldots, A_p\} \mid A_{(s,\lambda)} = (d_{(s,\lambda)}, C_{(s,\lambda)}) \wedge d_{(s,\lambda)} \in T_B \wedge \min(time(T_B)) + \Delta t \leq time(d_{(s,\lambda)}) < \max(time(T_B)) - \Delta t \tag{4}$$

All documents are pre-processed to remove numbers, URL's, email addresses, meaningless symbols, and formatting. Each term in the content C of the document is stemmed using the Porter stemmer [21]. The stemmed term index defined in Eq. (5) is created with the stemmed terms which appear in the document, and the number of times they appear.

$$C_{(s,\lambda)} = \{T_1, T_2, \ldots, T_q\} \mid T_{(s,\lambda,\varpi)} = \{S_{(s,\lambda,\varpi)}, SC_{(s,\lambda,\varpi)}\} \wedge SC_{(s,\lambda,\varpi)} = \#\{\forall S_{(s,\lambda,\varpi)} \in C_{(s,\lambda)}\} \tag{5}$$

3. Methodology

The methodology section is split into sections covering the Classification of the documents and the Training procedure.

3.1. Classification

In order to determine the accuracy of a classifier it is necessary to have specific measures of how the market reacts to news. To do so it is necessary to perform time series analysis on the data and classify each document according to how the market reacted shortly after its arrival. The return time series defined in Eq. (6) investigates the log returns over the period Δt for the stock. The return time series is one of the most interesting to investors as it demonstrates the amount of money which can be made. Abnormal returns should correlate more highly to the arrival of news because the market suddenly has more information to process.

The variable M in Eq. (7) defines the average number of trading minutes per month by using the average number of trading minutes per business day for the relevant country, and multiplying by the average number of trading days per month (20).

In Eq. (8) we define the mean for each trading minute in the return time series, by taking the mean value for the M trading minutes which preceded the start of the current trading day. In Eq. (9) we define the standard deviation for each trading minute in the return time series, by again using the M trading minutes which preceded the start of the current trading day. Note that if a stock was suspended from trading during the last 20 trading days for the stock exchange, we use the last 20 days which the stock traded on.

$$R_{(s,\Delta t)} = \left\{R_1, \ \ldots \ , R_m\right\} \ | \ \ R_{(s,t,\Delta t)} = \log\left(P_{(s,\ t)}\right) - \log\left(P_{(s,\ t-\Delta t)}\right) \tag{6}$$

$$M = 20 \times m \ | \ \left\{m_{US} = 390, \ \ m_{UK} = 510, \ \ m_{AU} = 360\right\} \tag{7}$$

$$\mu R_{(s,t,\Delta t)} = \frac{\sum_{j=t_0-M}^{t_0-1} R_{(s,j,\Delta t)}}{M} \ | \ \ t_0 = \min\left(\left\{\begin{matrix}\forall T_{(B,i)} \ | \ \ time\left(T_{(B,i)}\right) = \\ \min\left(time\left(T_B\right)\right) \ \wedge \ \ T_{(B,i)} \le t\end{matrix}\right\}\right) \tag{8}$$

$$\sigma R_{(s,t,\Delta t)} = \sqrt{\frac{\sum_{j=t_0-M}^{t_0-1}\left(R_{(s,j,\Delta t)} - \mu R_{(s,t,\Delta t)}\right)^2}{M}} \ | \ \ t_0 = \min\left(\left\{\begin{matrix}\forall T_{(B,i)} \ | \ \ time\left(T_{(B,i)}\right) \\ = \min\left(time\left(T_B\right)\right) \ \wedge \ \ T_{(B,i)} \le t\end{matrix}\right\}\right) \tag{9}$$

We classify the outcome O of each article in Eq. (4) using the definition in Eq. (10) in which we require that the return within τ minutes is equal or exceeds δ standard deviations from the mean function value.

$$O_{(s,\Delta t,\tau,\delta)} = \left\{O_1, O_2, \ \ldots \ , O_p\right\} \ | \ \ O_{(s,\Delta t,\tau,\delta,\lambda)} = \tag{10}$$

$$\left(\exists t \,|\, d_{(s,\lambda)} < t \le d_\lambda + \tau \ \wedge \ \left(\begin{matrix} R_{(s,t,\Delta t)} \ge \mu R_{(s,t,\Delta t)} + \delta \times \sigma R_{(s,t,\Delta t)} \\ \vee \ R_{(s,t,\Delta t)} \le \mu R_{(s,t,\Delta t)} - \delta \times \sigma R_{(s,t,\Delta t)}\end{matrix}\right) \ ? \ 1 \ : \ 0\right)$$

3.2. Training

The stocks for each country c are grouped together using Eq. (11) to form a large dataset of related stocks. Each document for each stock within each country is then classified using the return time series with the chosen parameters. For each country and value of δ we create 10 training and tests sets for the sake of robustness.

We create training sets by taking Φ documents at random, of which Ψ were classified as interesting, and the rest were not. The remaining documents are used as the test set.

A dictionary is created using Eq. (12) for each term which appears in at least one document for a stock in the training set. We store the term count TC, document count DC, interesting term count ITC, and interesting document count IDC for each term. The TC is the total number of times the given term appears in all documents in the training set. The DC is the total number of documents which contain the given term. The ITC is the total number of times the given term appears in all documents which are classified as interesting in the training set. The IDC is the total number of documents which are classified as interesting in the training set which contain the given term.

$$G_{(c)} = \{G_1, G_2, \ \dots \ , G_v\} \tag{11}$$

$$\begin{aligned} & X_{(c,\Delta t,\tau,\delta)} = \{X_1, X_2, \ \dots \ , X_w\} \\ & \mid \ X_{(c,\Delta t,\tau,\delta,\varpi)} = \{S_{(c,\Delta t,\tau,\delta,\varpi)}, TC_{(c,\Delta t,\tau,\delta,\varpi)}, DC_{(c,\Delta t,\tau,\delta,\varpi)}, ITC_{(c,\Delta t,\tau,\delta,\varpi)}, IDC_{(c,\Delta t,\tau,\delta,\varpi)}\} \\ & \quad \wedge TC_{(c,\Delta t,\tau,\delta,\varpi)} = \sum SC_{(s,\lambda,\varpi)} \ \mid \ s \in G_{(c)} \\ & \quad \wedge DC_{(c,\Delta t,\tau,\delta,\varpi)} = \#\{\forall C_{(s,\lambda)} \ \mid \ SC_{(s,\lambda,\varpi)} > 0 \ \wedge \ s \in G_{(c)}\} \\ & \quad \wedge ITC_{(c,\Delta t,\tau,\delta,\varpi)} = \sum SC_{(s,\lambda,\varpi)} \ \mid \ O_{(s,\Delta t,\tau,\delta,\lambda)} = 1 \ \wedge \ s \in G_{(c)} \\ & \quad \wedge IDC_{(c,\Delta t,\tau,\delta,\varpi)} = \#\{\forall C_{(s,\lambda)} \ \mid \ SC_{(s,\lambda,\varpi)} > 0 \ \wedge \ O_{(s,\Delta t,\tau,\delta,\lambda)} = 1 \ \wedge \ s \in G_{(c)}\} \end{aligned} \tag{12}$$

A sub-dictionary is formed by taking the top ψ terms based on a given term ranking algorithm. Firstly we chose the term frequency inverse document frequency (TFIDF) algorithm defined in Eq. (13). It is calculated by combining the product of the term frequency (first part of equation) with the inverse document frequency (log part of equation). Note the Φ in Eq. (13) is the number of documents in the training set. We sort the values in descending order such that terms that occur most frequently are chosen, as this is the order which generally gives the best results when querying documents.

$$TFIDF_{(c,\Delta t,\tau,\delta,\varpi)} = TC_{(c,\Delta t,\tau,\delta,\varpi)} \times \log_{10}\left(\frac{\Phi}{DC_{(c,\Delta t,\tau,\delta,\varpi)}}\right) \tag{13}$$

Secondly we chose the binary version of Quinlan's Gain ratio [22], as defined in Eq. (15). This algorithm selects terms which provide the most information, i.e. split the data between the classes most effectively. In Eq. (15) the $E(\Psi, \Phi)$ part of the equation is the entropy value for the ratio of interesting documents (Ψ) to documents (Φ) in the training set. The next part of the equation calculates the entropy value for the ratio of interesting documents to documents which contain the term, scaled by the

ratio of documents which contain the term. The last part of the equation calculates the entropy value for the ratio of uninteresting documents to documents which contain the term, scaled by the ratio of documents which don't contain the term.

$$E(n,N) = -\left(\frac{n}{N}\log_2\left(\frac{n}{N}\right) + \left(1-\frac{n}{N}\right)\log_2\left(1-\frac{n}{N}\right)\right) \quad | \quad n \le N \tag{14}$$

$$GAIN_{(c,\Delta t,\tau,\delta,\varpi)} = E(\Psi,\Phi) - \frac{DC_{(c,\Delta t,\tau,\delta,\varpi)}}{\Phi} \times E\left(IDC_{(c,\Delta t,\tau,\delta,\varpi)}, DC_{(c,\Delta t,\tau,\delta,\varpi)}\right) - \frac{\Phi - DC_{(c,\Delta t,\tau,\delta,\varpi)}}{\Phi} \times E\left(DC_{(c,\Delta t,\tau,\delta,\varpi)} - IDC_{(c,\Delta t,\tau,\delta,\varpi)}, DC_{(c,\Delta t,\tau,\delta,\varpi)}\right) \tag{15}$$

Finally we adapted Robertson and Spärk Jones's BM25 algorithm (Best Match) [23] to get the Average Document BM25 value (ADBM25) defined in Eq. (16), where k_1 and b are constants. The ADBM25 algorithm is the same as the BM25 algorithm if Φ were equal to 1, or in other words if there was only one document. The first part of the equation normalises the term frequency by taking into account the length of the document which contains the term and the average document length. This ensures that if a term occurs frequently in a very long document, it isn't given unwarranted significance. The log part of the equation normalises results by factoring in the number of interesting documents which contain the term (IDC), the number of documents which contain the term (DC) and the total number of interesting documents (Ψ) and documents (Φ). This favours terms which provide more information, i.e. split the two classes most efficiently.

$$ADBM\,25_{(c,\Delta t,\tau,\delta,\varpi)} = \frac{1}{\Phi}\sum_{d=1}^{\Phi} \frac{(k_1+1)\times TC_{(c,\Delta t,\tau,\delta,\varpi)}}{\left(k_1 \times \left((1-b) + b \times \frac{dl_{(d)}}{avdl}\right)\right) + TC_{(c,\Delta t,\tau,\delta,\varpi)}} \times \log \frac{\left(IDC_{(c,\Delta t,\tau,\delta,\varpi)} + 0.5\right) \times \left(\Phi - DC_{(c,\Delta t,\tau,\delta,\varpi)} - \Psi + IDC_{(c,\Delta t,\tau,\delta,\varpi)} + 0.5\right)}{\left(DC_{(c,\Delta t,\tau,\delta,\varpi)} - IDC_{(c,\Delta t,\tau,\delta,\varpi)} + 0.5\right) \times \left(\Psi - IDC_{(c,\Delta t,\tau,\delta,\varpi)} + 0.5\right)}$$
$$| \quad dl_{(\lambda)} = \sum \forall SC_{(s,\lambda,\varpi)} \quad \wedge \quad avdl = \frac{1}{\Phi}\sum_{\rho=1}^{\Phi} dl_{(\rho)} \tag{16}$$

A binary vector is created for each document in the training and test set where each entry specifies whether the given term (which is a member of the sub-dictionary) occurred in the document. These vectors are used to train and test Quinlan's C4.5 decision tree [22], and Vapnik's support vector machine (SVM) [24] using the SVM Light Classifier [25].

The C4.5 decision tree [22] classifies documents by building a tree where the root node is the term which produces the highest Gain value (Eq. (15)). Each leaf node which branches from the root, or from subsequent branches, chooses the combination of terms which produces the highest Gain value. This value is calculated by combining the terms and their values (i.e. contains the term or not) of every node above the current node and adding one extra term. Only documents which have the given term values (i.e. contain or don't contain the specified terms as required) are included in the Gain equation. This ensures that the extra terms are appended based on their ability to separate the remaining documents into the two classes.

Vapnik's support vector machine (SVM) [24] projects the terms and their values into higher dimensional space (e.g. one dimension per term). It produces a classifier by identifying the hyperplane which most effectively separates the two classes.

To measure the performance of the classifiers we use the overall accuracy of the classifier defined in Eq. (17). It is calculated by dividing the total number of correctly classified documents (True Positives and True Negatives) by the total number of documents in the test set (Φ).

$$Accuracy = \frac{\#TP+\#TN}{\Phi} \tag{17}$$

4. Results

Each document is classified using the return time series with $\Delta t=\tau$ equal to the specified number of minutes, and $\delta=4$. For all tests there are 1,000 documents in the training set (Φ), of which 500 are interesting (Ψ). We chose an equal split so as not to skew the results of the classifier. Furthermore we use $k_1=1$ and $b=0.5$ for the ADBM25 term ranking algorithm in Eq. (16). Finally we run tests with varying φ values (100,200,500,1000,2000,5000).

The characteristics of the datasets are shown in Table 1 where Docs are the total number of documents which were released to the given country during the times specified in Eq. (4). The Int. Docs are those which were categorised as causing an abnormal return. The ratio is the percentage of documents which are interesting.

Table 1. Characteristics of different datasets, with $\delta=4$, $\Delta t=\tau$ having the listed values, G consisting of all stocks for the given country (US, UK, Australia).

	US			UK			Australia		
Δt,τ (minutes)	**Docs**	**Int. Docs**	**Ratio**	**Docs**	**Int. Docs**	**Ratio**	**Docs**	**Int. Docs**	**Ratio**
5	133,019	2,414	1.81%	81,528	2,046	2.51%	33,165	933	2.81%
10	129,370	2,760	2.13%	80,245	2,487	3.10%	31,728	1,259	3.97%
15	124,616	2,539	2.04%	78,871	2,756	3.49%	30,455	1,388	4.56%
30	112,907	2,205	1.95%	74,664	2,851	3.82%	27,054	1,232	4.55%
45	100,245	1,710	1.71%	70,054	2,698	3.85%	23,910	1,030	4.31%
60	89,159	1,452	1.63%	65,230	2,503	3.84%	20,835	913	4.38%
90	68,056	973	1.43%	54,230	1,899	3.50%	14,588	560	3.84%

The ratio actually increases as the time window size is increased before reducing again. This is because τ allows more time for the market to react to the news and Δt isn't large enough to limit the number of documents which occur in the time period.

The results in Fig. 1 show the results of the best classifier for each time window size ($\Delta t=\tau$) and country. The details of the term ranking algorithm, the classifier, the number of terms (φ) which produced these results are included in Table 2. This also includes the standard deviation of the classifier as well as the maximum achieve with the given parameters, where the best results are bolded.

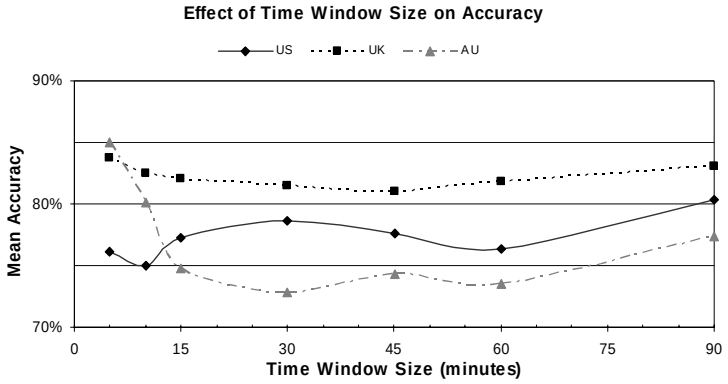

Fig. 1. The mean accuracy for the best classifier for each time window ($\Delta t=\tau$) and country**.**

Table 2. The characteristics of the best classifier for each time window ($\Delta t=\tau$) and country.

Country	Δt,τ	Term Ranking	Classifier	Terms (φ)	Accuracy	Maximum Accuracy
US	5	ADBM25	SVM	5,000	76.12±02.39%	78.95%
US	10	ADBM25	SVM	100	74.97±02.30%	79.73%
US	15	ADBM25	SVM	5,000	77.28±01.13%	79.02%
US	30	ADBM25	SVM	5,000	78.67±02.05%	83.32%
US	45	ADBM25	SVM	5,000	77.58±01.86%	80.30%
US	60	ADBM25	SVM	5,000	76.39±01.19%	78.75%
US	**90**	**GAIN**	**SVM**	**200**	**80.36±01.20%**	**82.55%**
UK	**5**	**GAIN**	**C4.5**	**100**	**83.72±01.33%**	**86.46%**
UK	10	GAIN	C4.5	100	82.46±01.08%	85.24%
UK	15	GAIN	C4.5	100	82.09±01.57%	84.52%
UK	30	GAIN	C4.5	100	81.47±00.93%	83.04%
UK	45	GAIN	C4.5	100	81.06±01.61%	84.92%
UK	60	GAIN	C4.5	100	81.84±00.92%	84.22%
UK	90	GAIN	C4.5	100	83.10±00.73%	84.80%
AU	**5**	**ADBM25**	**SVM**	**5,000**	**85.00±00.64%**	**86.60%**
AU	10	ADBM25	SVM	2,000	80.13±01.30%	82.75%
AU	15	ADBM25	C4.5	100	74.74±02.71%	78.98%
AU	30	ADBM25	C4.5	100	72.85±01.74%	76.56%
AU	45	ADBM25	C4.5	100	74.27±01.94%	77.73%
AU	60	ADBM25	C4.5	100	73.54±01.35%	75.40%
AU	90	ADBM25	C4.5	100	77.37±01.24%	80.81%

In the US it appears that investors react to some news within 5 minutes and tend to consistently react within 30 minutes of news. However the best results are achieved by allowing 90 minutes for the market to react to news. This indicates that investors in the US tend to pay more attention to market movement than to the release of news. This is because the US market is the largest stock market in the world and therefore has the most frequent trading. Therefore investors need to pay close attention to how other market participants are behaving in order to make their decisions. This leaves

them less time to read news and therefore probably only read news once they have noticed abnormal returns which they may capitalise on. Furthermore there is significantly more news released to the US market than the other two markets. Therefore investors are less likely to read many articles as they are likely to see most news as irrelevant unless it has a catchy headline.

Investors in the UK tend to react quickly and decisively to news within 5 minutes and continue to react in a similar fashion for long after. This indicates that investors in the UK not only pay close attention to news but they also consistently react in a similar fashion to news with similar content. The slight reduction in accuracy over time is an indication that market noise has introduced articles which themselves where not responsible for any change.

It appears that investors in the Australian market are less rational than those in the other two markets. This is because they react quickly and decisively to news within 5 minutes but the accuracy dramatically reduces when the time window is increased. This should be expected as the Australian market is significantly smaller than the other two markets and therefore has considerably less trading than the others. Therefore there can be long periods where there has been little or no trading which leads to a lower standard deviation of the return. However some investors, most likely large institutional investors, must pay close attention to news in order to react to the news consistently and quickly. The rest of the market however probably either has less or delayed access to public information and therefore don't tend to react in a similar fashion. Alternatively it could mean that investors in Australia are somewhat irrational as they don't consistently react in the same way to the same news.

Table 3. The characteristics of the best classifier with φ limited to 100 and 200 for each country and term ranking algorithm.

			TFIDF		**GAIN**		**ADBM25**	
Country	**Δt,τ**	**φ**	**Accuracy**	**Class**	**Accuracy**	**Class**	**Accuracy**	**Class**
US	90	100	73.45±01.15%	SVM	**78.24±10.07%**	SVM	75.56±01.68%	SVM
US	90	200	71.77±01.61%	SVM	**80.36±01.20%**	SVM	76.40±02.02%	SVM
UK	5	100	69.52±04.29%	SVM	**83.72±01.33%**	C4.5	75.11±01.94%	SVM
UK	5	200	68.50±04.05%	SVM	**79.67±02.03%**	C4.5	75.80±01.82%	SVM
AU	5	100	**83.74±01.06%**	SVM	81.50±11.32%	C4.5	82.28±00.71%	SVM
AU	5	200	**83.46±00.92%**	SVM	74.75±15.72%	SVM	82.70±00.81%	SVM

The results in Table 3 show the most accurate classifier for each country and term ranking algorithm with the best time window and a limit of 100 and 200 terms. The most accurate results are bolded. Combining the results with those in Table 2 it is clear that the SVM is the best classifier for the US and Australian markets, whilst the C4.5 classifier is more useful in the UK market. This suggests that longer rules are required to classify documents in the UK market as the SVM tends to ignore terms which individually have little impact, whilst the C4.5 classifier is more comprehensive. However when the number of terms (φ) is increased beyond 200 the SVM is better than the C4.5 classifier in the UK though produces worse results than for the 100 and 200 terms tests.

In the US and UK the best results are obtained using the Gain term ranking algorithm. This implies that the training sets are indicative of the entire dataset. However with time windows ($\Delta t=\tau$) less than 90 minutes it is necessary to include

more terms in the US and in that case the ADBM25 term ranking algorithm is better. This is because annual reports released in the US are very long whilst some analyst recommendations and similar news are very short. Therefore it is necessary to account for the length of the document when ranking the effect of the term.

In the Australian market the TFIDF term ranking algorithm yields the best results when the number of terms (φ) is limited. This indicates that the relevance of the term is less important in Australia. We can assume this because both the Gain and ADBM25 term ranking algorithms account for the relevance of the term, whilst TFIDF does not. This could imply that investors in Australia are less rational because they don't react consistently to similar news. Alternatively it could mean that news providers for the Australian market don't use the same terminology. Therefore there are many different ways to say the same thing which leads to difficulty in seeing the similarity between documents.

The Receiver Operating Characteristic (ROC) curves in Fig. 2 are for the classifier, term ranking algorithm, and time window combination which produced the most accurate classifier for each country. All results are clearly better than the line of no-discrimination though the number of true positives found is quite low. This is because it is necessary to choose a classifier with the least number of false positives, as there are considerably more uninteresting documents. However the results show that it is possible to predict whether news will cause abnormal returns.

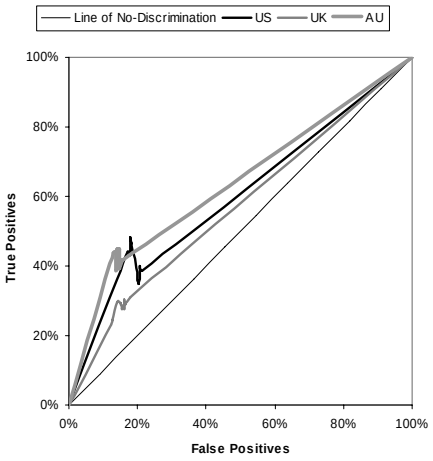

Fig. 2. The ROC curves of the best classifier for each country**.**

5. Conclusions

We have classified news based to abnormally large returns and then employed the SVM and C4.5 classifiers to forecast the short-term market reaction to news. We have also utilised the TFIDF, Gain, and a modified version of the BM25 term ranking algorithms to aid the decisions of the classifiers.

We have found that the Gain term ranking algorithm is superior in the US and UK markets. This implies that the training sets are representative of the entire dataset as the Gain term ranking algorithm chooses terms which have high information value within the training set. This means that investors in the US and UK appear to be more rational as they react the same way to similar documents.

The SVM classifier was discovered to have the best performance in the US and Australian markets whilst the C4.5 classifier was better for the UK market. This suggests that longer rules are required to classify documents in the UK market as the SVM tends to ignore terms which individually have little impact, whilst the C4.5 classifier is more comprehensive.

The fact that the Gain and ADBM25 term ranking algorithms are less effective in the Australian market when the number of terms (φ) is limited, indicates that Australian investors are somewhat irrational (i.e. make decisions without any new information). This is probably because the Australian market is substantially smaller than the other two markets and therefore there are less large institutional investors, who tend to pay close attention to the market. Alternatively it could mean that news providers in the Australian market don't use consistent terminology. This means there are many ways to say the same thing, and therefore the relevance of the document is less important.

We have found that the 90 minute time window yields the most accurate classifier in the US. This suggests that investors in the US tend to pay more attention to actual market behaviour than to the release of news. This is because the US stock market is the largest in the world and therefore has more frequent trading. Therefore investors must pay attention to trading and therefore they have less time to read news. Furthermore there is far more news released in the US than the other two markets. Therefore investors probably tend to ignore articles unless they have compelling headlines or they notice the market behaving differently.

The UK and Australian markets react quickly and decisively to some news within 5 minutes. This is most likely annual reports, or earnings announcements which investors were anxiously awaiting and therefore the documents have similar content. However the classification accuracy drops as the time window is increased before eventually rising again. This effect is significantly more noticeable in the Australian market than the UK. This implies that investors in the UK pay close attention to news and are rational as they consistently react in the same way to news with similar content. In Australia however it appears that the large institutional investors are responsible for the decisive reaction within 5 minutes. This is because they have undelayed access to most news and the staff to read it. However other investors either have delayed access or are slightly irrational as they don't consistently react to news with similar content.

These results are promising and it would appear that it is cost-effective to develop a system which highlights interesting news for investors based on its content.

References

1. Oberlechner, T. and S. Hocking: Information Sources, News, and Rumours in Financial Markets: Insights into the Foreign Exchange Market. Journal of Economic Psychology, Vol. 25. (2004) 407-424.
2. Cutler, D.M., J.M. Poterba, and L.H. Summers: What Moves Stock Prices? Journal of Portfolio Management, Vol. 15. (1989) 4-12.
3. Goodhart, C.A.E., News and the foreign exchange market. In Proceedings of Manchester Statistical Society. 1989: p. 1-79.
4. Goodhart, C.A.E., S.G. Hall, S.G.B. Henry, and B. Pesaran: News Effects in a High-Frequency Model of the Sterling-Dollar Exchange Rate. Journal of Applied Econometrics, Vol. 8. (1993) 1-13.
5. Mitchell, M.L. and J.H. Mulherin: The Impact of Public Information on the Stock Market. Journal of Finance, Vol. 49. (1994) 923-50.
6. Melvin, M. and X. Yin: Public Information Arrival, Exchange Rate Volatility, and Quote Frequency. Economic Journal, Vol. 110. (2000) 644-661.
7. Mittermayer, M.-A., Forecasting Intraday Stock Price Trends with Text Mining Techniques. In Proceedings of 37th Annual Hawaii International Conference on System Sciences (HICSS'04), Big Island, Hawaii. 2004: p. 30064b.
8. Ederington, L.H. and J.H. Lee: How markets process information: News releases and volatility. Journal of Finance, Vol. 48. (1993) 1161-1191.
9. Ederington, L.H. and J.H. Lee: The short-run dynamics of the price adjustment to new information. Journal of Financial & Quantitative Analysis, Vol. 30. (1995) 117-134.
10. Ederington, L.H. and J.H. Lee: Intraday Volatility in Interest-Rate and Foreign-Exchange Markets: ARCH, Announcement, and Seasonality Effects. Journal of Futures Markets, Vol. 21. (2001) 517-552.
11. Almeida, A., C.A.E. Goodhart, and R. Payne: The Effects of Macroeconomic News on High Frequency Exchange Rate Behavior. Journal of Financial & Quantitative Analysis, Vol. 33. (1998) 383-408.
12. Graham, M., J. Nikkinen, and P. Sahlstrom: Relative Importance of Scheduled Macroeconomic News for Stock Market Investors. Journal of Economics and Finance, Vol. 27. (2003) 153-165.
13. Kim, S.-J., M.D. McKenzie, and R.W. Faff: Macroeconomic News Announcements and the Role of Expectations: Evidence for US Bond, Stock and Foreign Exchange Markets. Journal of Multinational Financial Management, Vol. 14. (2004) 217-232.
14. Nofsinger, J.R. and B. Prucyk: Option volume and volatility response to scheduled economic news releases. Journal of Futures Markets, Vol. 23. (2003) 315-345.
15. Hong, H., T. Lim, and J.C. Stein: Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. Journal of Finance, Vol. 55. (2000) 265-95.
16. Womack, K.L.: Do Brokerage Analysts' Recommendations Have Investment Value? Journal of Finance, Vol. 51. (1996) 137-67.
17. Michaely, R. and K.L. Womack: Conflict of Interest and the Credibility of Underwriter Analyst Recommendations. Review of Financial Studies, Vol. 12. (1999) 653-86.
18. Roll, R.: Orange Juice and Weather. American Economic Review, Vol. 74. (1984) 861-80.
19. Wuthrich, B., D. Permunetilleke, S. Leung, V. Cho, J. Zhang, and W. Lam, Daily Stock Market Forecast from Textual Web Data. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. 1998: p. 2720-2725.
20. Fung, G.P.C., J.X. Yu, and L. Wai, Stock Prediction: Integrating Text Mining Approach using Real-Time News. In Proceedings of IEEE International Conference on Computational Intelligence for Financial Engineering, Hong Kong. 2003: p. 395-402.
21. Porter, M.F.: An Algorithm for Suffix Striping. Automated Library and Information Systems, Vol. 14. (1980) 130-137.

22. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1993).
23. Robertson, S. and K. Spärck Jones: Simple, Proven Approaches to Text Retrieval. University of Cambridge Computer Laboratory Technical Report no. 356, Vol. (2006).
24. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag (1999).
25. Joachims, T.: SVM Light Classifier, (2007). Available: http://svmlight.joachims.org/.