Discovering Intraday Market Risk Exposures in Unstructured Data Sources: The Case of Corporate Disclosures

Sven S. Groth
E-Finance Lab Frankfurt
sgroth@wiwi.uni-frankfurt.de

Jan Muntermann
Goethe-University Frankfurt
muntermann@wiwi.uni-frankfurt.de

Abstract

Capital markets react promptly and significantly to critical events that have not been anticipated by market participants. Prominent examples of such market behaviour which risk management activities turn their attention to became evident during the last few months. Today, classic risk management tools perform complex calculations on the basis of structured data sources such as historical price series. In contrast, unstructured data sources, such as corporate disclosures, are widely disregarded. However, such data that has been released newly could carry information that is not reflected in the structured data available in such a situation. We aim to utilize such unstructured data sources and present a text mining-based approach to detect intraday market risk exposures that result from the event of newly published information. Our results provide evidence that unstructured data contains valuable information to discover intraday risk exposures promptly, i.e. represent valuable data sources in this context.

1. Introduction

Financial modelling of market risks, i.e. the management of losses due to movements in financial market prices, has been a subject of research for decades. Today, traditional financial risk management tools such as Value-at-Risk make use of quantitative data being stored in structured databases [21]. While the approaches to analyze structured data such as historical price series have been continuously improved, little attention has been paid to the analysis of unstructured qualitative data in this context in the past. Especially, when assessing intraday market risk exposures that result from market events such as critical corporate disclosures that were not anticipated by market participants, there exists little to no quantitative data that could be analyzed in such situations. In contrast, there exists qualitative data (i.e. the disclosures content) representing a potential source of information that is not sufficiently taken into consideration by traditional risk management tools.

In this paper, we present a text mining-based approach that aims at analyzing such unstructured (and so far widely neglected) data sources in order to detect relevant patterns. These patterns should then provide a basis to identify and mitigate intraday market risk exposures. To our best knowledge, the application of text mining techniques, i.e. the analysis of qualitative textual data in the area of financial risk management is a novel approach to support decision making in this context. Our analysis is based on a collection of corporate disclosures for which empirical research provides evidence that they represent a significant source of market risk since significant price reactions have been observed subsequent to their publication. We analyze the disclosures’ content utilizing state-of-the-art text mining techniques to identify those events that will result in most extraordinary market reaction. These reactions are measured by volatility increases of the corresponding intraday stock returns that we observed subsequent to the disclosures’ publication dates. Being able to detect patterns that drive these volatilities, i.e. to identify most critical market events, should enable decision makers to take measures to mitigate the corresponding risk exposure.

Our paper is organized as follows: Next, we provide a literature review on relevant financial and text mining research. Then, we present our approach to discover intraday market risk exposures in unstructured data sources (corporate disclosures). The following two sections illustrate how our approach can be evaluated with regard to its classification performance and its capability to mitigate intraday market risks. Therefore, we first present an evaluation on the basis of “classic” evaluation measures. Then, we second present a “simulation-based” evaluation approach. Here, we present a financial vehicle (straddle option) that provides the basis for a newly developed domain-specific evaluation metric. The evaluation results provide strong evidence that our text mining approach is capable to identify critical market events on the basis
of qualitative disclosure contents. Finally, we summarize and conclude that our findings encourage the incorporation of unstructured data sources into novel financial risk management tools.

2. Literature review

2.1. Capital market research and risk management

The effect of unforeseen events on capital markets has been explored in empirical financial research for decades. This research stream is based on the semi-strong form of the efficient market hypothesis that implies that share prices promptly adjust to new publicly available information [9]. In order to address the question what kind of events are of relevance, so-called event studies are conducted to analyze price effect magnitudes and the speed at which prices fully reflect the new information available. Empirical evidence suggests that significant abnormal price behaviour can be observed subsequent to the publication of corporate disclosures, especially for disclosures that were published due to regulatory legislation. Significant abnormal market behaviour has been observed in this context worldwide in the past [25, 30]. With prompt price reactions that can exceed multiple standard deviations of historical price movements, such corporate disclosures represent an event type with significant risk exposure.

Today, such market risks are modelled utilizing modern information systems analyzing huge amounts of quantitative data. Investment companies, for example, run complex and time consuming simulations in order to assess current risk positions. Current risk positions that are assessed on the basis of historical data (e.g. when conducting historical simulations) or repeated random sampling (e.g. Monte Carlo simulation) are usually measured by a Value-at-Risk measure that estimates the risk of loss on a given portfolio of financial assets at a certain confidence interval [19]. Due to this confidence interval, it does not provide support for managing extraordinary losses and since the underlying calculations are based on historical or artificial datasets, intraday losses that could result from critical market events (event risks) are not covered [1]. Another established risk management approach that has a focus on events that could result in extraordinary losses is stress testing. Also utilizing historical or artificial quantitative data, stress testing aims at supporting managers to assess financial consequences of critical market behaviour, i.e. to conduct what-if-analyses for such scenarios [18]. Central objective of these approaches is to identify and reduce hypothetical risk exposures but not to discover actual intraday risk exposures that result from given but unforeseen critical events.

2.2. Text mining research

Unexpected and / or simply plain volatility undoubtedly constitutes a source of potential risk. The fact that there is already a multitude of research on the measurement concepts of volatility underpins this perception [26]. Nonetheless, there exists only little – depending on the definition, if any at all – research that aims at utilizing unstructured data sources in the context of risk management by applying text mining techniques. Even though the data / text mining approaches by Schulz et al. [29] and Thomas [31] are classified as volatility forecasting systems by Mittermayer [24], we believe that this is not true without limitations: Schulz et al. [29] try to forecast abnormal stock returns; not volatility. The rule-based hand-crafted classifier by Thomas [31] does not reflect our understanding of text mining (see below, [13]).

Another stream of literature that is similar to our proposed risk mitigation text mining approach can be found in the literature on forecasting stock price movements. Wuthrich et al. [36] present one of the first applications of text mining techniques addressing financial forecasting problems. Later works focus on further aspects such as intraday events, applying new data mining techniques, varying the forecasting object, focusing on certain news types, or presenting novel evaluation methods [11, 23, 29].

3. Discovering intraday market risk exposures with text mining techniques

3.1. Study setup and dataset description

Having conducted a literature review on existing event study research, we know that the publication of corporate disclosures is oftentimes followed by significant abnormal stock returns. These significant (unexpected) stock movements represent a potential source of market risk. Aiming at reducing the associated risk exposure, we propose a text mining approach that will make predictions of future volatility levels (Figure 1). The predictions will be made for the time period directly following the publication of corporate disclosures. The approach is designed to identify those corporate disclosures that are associated with highest (abnormal) extra market movements. The introduced text mining approach is also applied on a real-world dataset by means of an empirical study. An evaluation of classification results is undertaken in two ways: First, we calculate “classic”
data mining evaluation metrics such as accuracy, recall, or precision [13]. As these measures, however, do not allow a “final” (statistical) conclusion about the suitability of the proposed text mining approach, an additional “simulation-based” performance evaluation is conducted.

![Figure 1. Study setup](image)

Our dataset contains all corporate disclosures that were published according to article 15 of the German Securities Trading Act between 2003-08-01 and 2005-07-31. With a focus on intraday market risks, we collected 423 disclosures that were published during stock exchange trading hours.

For each disclosure, the disclosure content, the publication date (exact to the minute) and the stock exchange symbol of the company that has initiated the disclosure’s publication has been collected. Given these stock symbols, the corresponding intraday (high-frequency) stock prices were collected from Frankfurt Stock Exchange for publication dates plus a period of ten days prior to these dates.

### 3.2. Defining intraday market risk

As our objective can be found with the identification of intraday market risk, we require valid predictions of future volatility. Moreover, those events that exhibit especially high volatility levels are more interesting to us than those events that exhibit especially low volatility levels. This is due to the fact that we aim at identifying situations that could create extraordinary risk exposure.

To model intraday market risk exposure being associated with critical market events such as the publication of corporate disclosures, we apply a model of risk that is based on an adjusted standard deviation ($aSTDEV$) of the rate of return for day $T$ during time interval $[t_1, t_2]$ (Formula 1). Thereby, the time period between $t_1$ and $t_2$ shall be denoted as $\tau$ (in minutes).

\[ aSTDEV = \sigma_{\tau, [t_1, t_2]} = \sqrt{f(\tau)} \cdot \sigma_{\tau, [t_1, t_2]} \]  

(Formula 1)

For reasons of better comparison of calculations with time periods of different length ($\tau$), the standard deviation is adjusted, i.e. annualized, by multiplying the variance (Formula 3) with factor $f(\tau)$ (Formula 2). It is thereby assumed that a year consists of 250 trading days. The number of trading hours per day is given by term $h$. The length of trading hours ($h$) needs to be modelled as a variable because it changed during the observation period from 11 hours to 8.5 hours.

\[ f(\tau) = 250 \cdot h \left( \frac{60}{\tau} \right) \]  

(Formula 2)

Fixed one-minute returns $x_i$ at day $T$ during time interval $[t_1, t_2]$ serve as input into the variance calculation (Formula 3). For each minute during interval $\tau$ the last tick price, i.e. the one closest to the minute under consideration, is used as $x_i$. The number of return observations $n$ is equivalent to $\tau$ (in minutes). $\mu$ constitutes the sample mean.

\[ \sigma_{\tau, [t_1, t_2]}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \]  

(Formula 3)

On the basis of the risk measure $aSTDEV$, we define an abnormal risk measure $ARisk_T$, (Formula 4) that adjusts the risk exposure following the event date $T_0$ by a risk exposure that has been observed before $T_0$, i.e. when no critical market event has occurred. Basically, event $aSTDEV$ is adjusted for the average $aSTDEV$ during previous $N$ days (in that particular stock). As literature suggests that there is usually an intraday volatility U-shape [34], i.e. high volatility at the open and close of the trading day, previous $N$ days’ $aSTDEV$ is calculated for the same time period $[t_0, f_0 + 1]$ as event $aSTDEV$.

\[ ARisk_T = \sigma_{t_0, [t_0, f_0 + 1]} - \frac{1}{N} \sum_{j=1}^{N} \sigma_{t_0, [t_0, f_0 + 1]} \]  

(Formula 4)

In order to evaluate whether or not above introduced news type “corporate disclosure” is actually associated with a high degree of uncertainty (risk), we calculate $ARisk_T$ for above introduced dataset. If both the mean ($t$-test) and the median (Wilcoxon signed rank test) of $ARisk_T$ values turn out to be significantly different from zero, the chosen event type “regulatory-driven corporate disclosures” constitutes a critical market event for which significant intraday market risk exposure can be expected.

We decided to calculate $ARisk_T$ for the following periods $\tau = 15$ minutes and 30 minutes for the following reasons: First, previous intraday event studies provide evidence that most significant abnormal market movements can be observed for these short periods of time subsequent to the publication of relevant information [27, 30]. Second, shorter periods than
\( \tau = 15 \) minutes involve problems regarding the applied risk measure \( \text{aSTDEV} \). For example, Andersen et al. [2] calculate volatilities for \( \tau = 5 \) minutes and encounter serious problems with the “bid-ask bounce”. Moreover, other authors argue that an interval as low as \( \tau = 15 \) minutes is sufficient for intraday volatility measurement [3].

Descriptive statistics and empirical test results for the 423 corporate disclosures \((N = 9)\) can be found in Table 1. For both the tests statistics the null hypotheses (also for both samples) can be rejected at the .01 level of significance. We conclude that corporate disclosures constitute critical market events representing sources of substantial market risk.

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics and test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{ARisk}<em>{\tau=15} ) &amp; ( \text{ARisk}</em>{\tau=30} )</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>75% Quartile</td>
</tr>
<tr>
<td>( t )-value</td>
</tr>
<tr>
<td>Wilcoxon signed rank value</td>
</tr>
</tbody>
</table>

*** indicates significance at the 1%-level.

The respective empirical \( \text{ARisk} \) distributions are illustrated in Figure 2. It can be observed that the distributions comprise of long tails on their right side. These long tails constitute most volatility-enhancing corporate disclosures. As the long tails are separated quite well from the rest of the empirical distribution by the 75% quartiles, we use these as input (labelling) to our text mining approach.

![Figure 2. Empirical ARisk distributions](image)

### 3.4. Text pre-processing

Analogue to Hotho et al. [13] we interpret text mining as “the application of algorithms and methods from the fields of machine learning and statistics to texts with the goal of finding useful patterns”. But as “traditional” machine learning methods are not able to cope with plain texts, the textual documents in the document collection need to be transformed into a numeric representation. This is done by employing – or at least taking account of – the three pre-processing steps Feature Extraction, Feature Selection, and Feature Representation [6].

During Feature Extraction a dictionary of words and phrases that describes the document collection adequately is generated. Thereby, a simple StringTokenizer [35] splits up the whole text into individual units, i.e. words. In order to eliminate “noise”, i.e. words with little meaning but frequent appearance, from the feature set a (German) stop word list and a threshold on the number of documents each token occurs in, has been made use of. Moreover, different grammatical forms of a word are mapped to a common stem by applying the Porter Stemmer [28]. Even though the Porter Stemmer [28] was not originally developed for the German language, our own pre-tests (not shown here) revealed that – everything else being equal – it performs better than an alternative simple GermanStemmer [35].

Within text mining the feature set can become very large, i.e. may contain some ten-thousand features. That is the reason why those features that contain few or relatively less information may be deleted during Feature Selection. We, however, do not apply this pre-processing step for the following reasons: First, we have sufficient computing resources. Second, Forman [10] finds that the available feature selection methods do not necessarily “perform better than using all features available”. Third, Joachims [17] provides further evidence that even features ranked lowest still
contain considerable information – according to their (binary) information gain – and are somewhat relevant. Fourth and most importantly, the applied data mining technique SVM is “nearly independent of the dimensionality of the feature space” [13] and is therefore not expected to suffer from the “curse of dimensionality”.

Finally, during Feature Representation each document is represented by previously extracted and selected number of features. The respective feature weightings in the document-feature matrix \( W \) is given by tfidf. \( Tf \) denotes the term frequency of a feature in a document and \( df \) denotes the number of documents the feature appears in [20].

### 3.5. Classification technique

Both our own pre-tests (not shown here) and comparative empirical studies [17, 38] provide evidence that the classification performance of SVM is superior to both parametric data mining techniques, e.g. Naïve Bayes, and non-parametric data mining techniques, e.g. k-Nearest Neighbour or Neural Networks. Moreover, as already stated above, SVM “is usually less vulnerable to the over-fitting problem [and] the solution of SVM is always unique and globally optimal”. That is the reason why we decided SVM to be the method of choice in this paper.

SVM was first introduced by Vapnik [33] for solving two-class recognition problems. The basic idea is to find a decision surface that maximizes the margin between data points, i.e. classes, by means of structural risk minimization. In case of originally non-separable data points, the original data vectors may be mapped to higher dimensional space to achieve linear separability again [38]. In order to reduce complexity, however, kernels, i.e. functions in lower dimensional space that exhibit similar behaviour as the original functions in higher dimensional space, are applied. We are making use of a linear kernel because Hsu et al. [14] provide evidence that, compared to other kernels, a linear kernel seems sufficient whenever the number of features is exceptionally large. The SVM implementation used in this study is the one provided by Chang and Lin [7], i.e. LIBSVM.

### 3.6. Post-processing

Besides “crisp” classification results the applied method SVM also delivers a “soft” probability / confidence value. It can be interpreted as a “guarantee of the learner that the corresponding crisp prediction is actually the true label” [22]. Documents are assigned to the classes “positive” or “negative” depending on whether the probability value is above or below a certain (learned) threshold. Varying the respective threshold will produce different points in the Receiver Operating Characteristic (ROC) space [5] and illustrates the inherent trade-off between precision and recall. The variation of thresholds may, however, also be applied as a “post-processing” step to account for imbalanced datasets or unequal classification costs [37]. We are confronted with both an imbalanced dataset and unequal classification costs:

As labelling is conducted on the 75% quartile, the class “positive” – by definition – contains 25% of all disclosures and the class “negative” contains the remaining 75% of all disclosures. By simply classifying each document as “negative”, one would already achieve a comparatively high accuracy of 75%.

Within risk management, we are mostly interested in those events that entail especially high risk exposure. Therefore, a good predictive performance for the class “negative” is not as useful to us as the same predictive performance for the class “positive”. In other words, our misclassification costs for the class “positive” are higher than for the class “negative”.

To overcome these problems, cost-sensitive learning is needed [16]. It shall, however, be clarified that cost-sensitive learning is not only a possible solution to unequal classification costs, but is “a good solution to the class imbalance problem” [39], too. It is against this background that we also make use of a ThresholdFinder [22] that uses the SVM confidence values to turn the SVM into a cost-sensitive learner.

### 4. “Classic” model evaluation

#### 4.1. “Classic” evaluation setup

Model evaluation is undertaken by means of k-fold \((k = 10)\) cross validation. Sub-samples are “stratified”, i.e. class distributions remain almost the same after sampling. Each \((1/k)\) test sub-sample contingency table is aggregated to create a global contingency table (micro averaging). The global contingency table is used to calculate the “classic” performance measures accuracy, recall, and precision [13]. As there is a trade-off between recall and precision, the figures shall not be assessed in isolation. We therefore additionally calculate the \(F_1\) measure by van Rijsbergen [32], where recall and precision is given equal weight (see Formula 5).

\[
F_1 = (2 \cdot \text{recall} \cdot \text{precision}) / (\text{recall} + \text{precision})
\]  

#### 4.2. “Classic” evaluation results

Classification results for the class “positive” can be found in Table 2 for \(\tau = 15\) and Table 3 for \(\tau = 30\).
These results provide the following insights: High (misclassification) costs of the class “positive”, e.g. 0.9, result in a high precision figure and a low recall figure. In other words, there are only very few corporate disclosures assigned to the class “positive” (i.e. low recall). Nonetheless, the majority of those disclosures that were assigned to the class “positive” actually belong there (i.e. high precision). This is due to the fact that disclosures are merely assigned to the high-risk class “positive” if the respective SVM confidence value is quite high, i.e. greater than the (learned) threshold.

Table 2. SVM (τ = 15) classification results *

<table>
<thead>
<tr>
<th>Misclassification cost for class</th>
<th>Accuracy (in %)</th>
<th>Recall (in %)</th>
<th>Precision F1 (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>77.07</td>
<td>8.49</td>
</tr>
<tr>
<td>0.3</td>
<td>0.9</td>
<td>78.49</td>
<td>31.13</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>77.30</td>
<td>47.17</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
<td>69.50</td>
<td>78.30</td>
</tr>
<tr>
<td>0.9</td>
<td>0.5</td>
<td>61.94</td>
<td>88.68</td>
</tr>
<tr>
<td>0.9</td>
<td>0.3</td>
<td>52.25</td>
<td>96.23</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
<td>42.79</td>
<td>100.00</td>
</tr>
</tbody>
</table>

* Results are given for the class of most interest, i.e. “positive”.

Table 3. SVM (τ = 30) classification results *

<table>
<thead>
<tr>
<th>Misclassification cost for class</th>
<th>Accuracy (in %)</th>
<th>Recall (in %)</th>
<th>Precision F1 (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>77.54</td>
<td>11.32</td>
</tr>
<tr>
<td>0.3</td>
<td>0.9</td>
<td>78.49</td>
<td>19.81</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>76.83</td>
<td>49.06</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
<td>71.87</td>
<td>66.98</td>
</tr>
<tr>
<td>0.9</td>
<td>0.5</td>
<td>51.30</td>
<td>94.34</td>
</tr>
<tr>
<td>0.9</td>
<td>0.3</td>
<td>44.92</td>
<td>98.11</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
<td>39.72</td>
<td>100.00</td>
</tr>
</tbody>
</table>

* Results are given for the class of most interest, i.e. “positive”.

The precision figures of 100% (τ = 15) and 92.31% (τ = 30) respectively provide strong evidence that the proposed text mining approach is capable of precisely identifying some of the most relevant high-risk entailing disclosures. Nevertheless, the high precision figure comes at the cost of low recall: In other words, we merely capture a small number of relevant high-risk disclosures. Therefore, one might be willing to accept a certain number of “false positives” to increase the number of “caught” high-risk disclosures. At the end, it is up to the investor to decide upon his / her individual threshold.

The accuracy is – at least in those cases that are most relevant to risk management applications – above the 75% guessing equivalent benchmark. It shall, however, be noted that this simple benchmark would assign each disclosure to the class “negative”. But as we are mostly interested in a good classification performance of the class “positive”, the figure accuracy in general and the benchmark in particular are not seen as good model evaluation methods in this context. We merely show them for reasons of completeness.

To conclude, depending on the respective threshold (i.e. misclassification cost) the proposed text mining approach seems to meet expectations. The “classic” model evaluation methods, however, do not allow us to draw “final” (statistical) conclusions. Moreover, we are not able to assess where the rightly assigned disclosures (“true positives”) are located in the high-risk quartile; i.e. at what part of the “fat tail” these are located. It follows that we need to conduct further evaluations that are able to grasp underlying finance-domain particularities.

5. “Simulation-based” evaluation

As seen in the previous section, the applied “classic” evaluation methods do not allow drawing a final conclusion about the use of text mining techniques to discover risk. Therefore, we further introduce a domain-specific “simulation-based” evaluation metric. While the domain specific “simulation-based” evaluation of stock movement predictions is straightforward and has already been applied many times in this context with an increasing degree of sophistication [11, 23], the domain specific “simulation-based” evaluation of volatility forecasts has – to our knowledge – not yet been applied in this context and has rarely been applied in other financial domains. Therefore, we are the first to develop and apply such an evaluation vehicle in this context.

5.1. Evaluation vehicle: straddle option

A (long) straddle option is a combination of both a (long) call option and a (long) put option with equivalent strike prices (see Figure 3). A long position in a straddle option should be built up if one expects stock prices to move significantly, but the direction of movement is unclear. In other words, a long straddle is a bet on increased volatility in the future. In order to be profitable, the stock price movement needs to be larger than the premiums (i.e. “c minus b”; “b minus a”) paid for both the call option and the put option. While the loss of a long straddle option is limited to the premiums paid, the potential profit is not capped. It follows that those stocks with highest unexpected / abnormal volatility, i.e. those that is not priced into option premiums, will also most likely reveal highest profits.
As the probability of profitable options (i.e. being heavily in-the-money) increases with higher volatility levels, volatility (risk) is a key determinant of option premiums. Figure 4 illustrates the sensitivity of the straddle option premium to two different inputs into the option pricing model [4], i.e. volatility and the risk-free rate. It can be seen that the variation of the risk-free rate between 0.5% and 6.0%, compared to the volatility input, has essentially no impact on the option premium.

To conclude, given that the volatility in this case is the most important input factor to option valuation and a key determinant of (straddle) option profitability, we believe that the “simulation-based” evaluation of a “long straddle strategy” constitutes a suitable additional tool to assess the suitability of above introduced text mining approach.

Sensitivity Analysis Example: Holding period = 30 minutes; asset price = 100; strike price = 100.

5.2. Simulation setup

The simulation is conducted on the same $k$-fold cross validation sub-samples as above analyses. In other words, it is ensured that the learned classification models are applied on independent test datasets. The remaining text mining steps such as labeling, preprocessing, or classification are equivalent to above approach, too.

Applying the SVM-model to the test dataset, each corporate disclosure in the test dataset is assigned to either the class “positive” or “negative”. If a corporate disclosure is assigned to the class “positive”, it is expected that it will entail a high abnormal volatility level after publication. Given that the proposed text mining approach is suitable for the underlying classification task, the (“positive”) disclosures will belong to the 25% most volatility-entailing news items. If this is the case, a long position in a straddle option is built up to profit from the (expected) high abnormal volatility. At option expiration ($t_0 + \tau$) either the call option or the put option are exercised. Exercising either one of the options is not undertaken if the strike price is equivalent to the stock price at expiration.

As we are mostly interested in finding the riskiest events, i.e. those belonging to the 25% most volatility-entailing corporate discoures, the respective evaluation investment strategy needs to be modelled accordingly. Therefore, a straddle “long-only” investment strategy is proposed. In other words, whenever a corporate disclosure is assigned to the class “negative” no action is undertaken.

5.3. Hypothetical option market

In order to calculate the premiums that need to be paid at event date ($T_0$) to build up a long position in a straddle option, a hypothetical option market is developed [8]. In this option market it is possible to simultaneously build up a long put and a long call position with equivalent strike prices. Moreover, the options mature at $T_0$, $T_0 + \tau$.

The size of option premiums is calculated using the Back-Scholes option pricing model [4]. Formula 6, 7, and 8 exemplarily show the calculation of a call option value $C_0$. Both the call option value $C_0$ and the put option value $P_0$ make up the straddle premium. The
exercise price $X$ is equivalent to the stock price $S_0$ at event time $(t_0)$. The annualized one-week EURIBOR (Euro Interbank Offered Rate) at event date $(T_0)$ was taken as input for the risk-free interest rate $r$. Time to maturity $M$ is given by the annualized $\tau$. The Black-Sholes option pricing model has already previously been applied in a different context on an intraday basis, i.e. with very short maturities [12].

$$C_0 = S_0 N(d_1) - X e^{-r \tau} N(d_2)$$

(6)

where

$$d_1 = \frac{\ln \left( \frac{S_0}{X} \right) + \left( r + \frac{\sigma^2}{2} \right) \tau}{\sigma \sqrt{\tau}}$$

(7)

$$d_2 = d_1 - \sigma \sqrt{\tau}$$

(8)

$C_0$ Current call option value  
$S_0$ Current stock price (at event time $t_0$)  
$N(d)$ Probability that a random draw from a standard normal distribution will be less than $d$  
$X$ Exercise price  
$r$ Risk-free interest rate (annualized)  
$M$ Time to maturity (annualized $\tau$)

As already stated above, the most important input factor to option valuation is volatility, i.e. $aSTDEV$. Above $ARisk$, results show that the chosen event type is – in most cases – followed by abnormal volatility levels. We believe that the market is aware of this fact. Therefore the options on the hypothetical market are priced accordingly, too. In other words, the hypothetical SVM-investor needs to pay high premiums to build up a long straddle: Each option is priced at $QUARTILE(aSTDEV, \tau)$. Thereby, $QUARTILE(aSTDEV, \tau)$ is defined as the 75% quartile of all documents’ $aSTDEV$ during the $\tau = 15$ and 30 minutes following the disclosure publication.

It shall be noted that labelling is conducted on an abnormal risk measure ($ARisk$) and the option premium calculation is based on a “normal” risk measure ($aSTDEV$). Both “classic” and “simulation-based” evaluation results for $aSTDEV$-labelling (not shown here), however, basically provide the same evidence as below. Furthermore, it shall also be taken into account that the proposed “simulation-based” approach merely constitutes an additional text mining evaluation metric. We do not pretend it to be a viable investment strategy per se. Investors would, for example, have a hard time to find options that bear above described characteristics.

5.4. “Simulation-based” evaluation results

The simulation setup illustrated in Figure 5 produces investment decisions for the test dataset based on the SVM-model. The respective SVM return population is termed $R_{SVM(cost \ “negative”; \ cost \ “positive”)}$ depending on the applied misclassification cost input.

The SVM return population is compared to a benchmark strategy whose return population is denoted as $R_{LONG}$. Hereby, a long straddle position is built up for each event in the test dataset irrespective of the content of the 423 corporate disclosures. We decided to use an “all-long” strategy as benchmark because it is closest to our objective of finding risk-entailing events. After all, above study has shown that corporate disclosures constitute critical market events. Furthermore, this benchmark strategy produced highest mean returns compared to – for instance – an all-short strategy. Nonetheless, it shall again be taken into account that the investment strategy merely constitutes another evaluation metric and shall not serve as motivation to actually follow such a strategy.

Descriptive statistics of return populations $R_{SVM(cost \ “negative”; \ cost \ “positive”)}$ and $R_{LONG}$ can be found in Table 4 (for $\tau = 15$) and Table 5 (for $\tau = 30$).

**Table 4. Simulation ($\tau = 15$) descriptive results**

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Misclassification cost for class</th>
<th>Population size</th>
<th>Mean in %</th>
<th>$aSTDEV$ in %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>neg. pos.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>0.1 0.9</td>
<td>9</td>
<td>7.3411</td>
<td>5.4979</td>
</tr>
<tr>
<td>SVM</td>
<td>0.3 0.9</td>
<td>51</td>
<td>5.4870</td>
<td>9.2536</td>
</tr>
<tr>
<td>SVM</td>
<td>0.5 0.9</td>
<td>90</td>
<td>4.3599</td>
<td>7.6680</td>
</tr>
<tr>
<td>SVM</td>
<td>0.9 0.9</td>
<td>189</td>
<td>3.2513</td>
<td>6.7671</td>
</tr>
<tr>
<td>SVM</td>
<td>0.9 0.1</td>
<td>348</td>
<td>2.3729</td>
<td>6.1789</td>
</tr>
<tr>
<td>LONG</td>
<td>- -</td>
<td>423</td>
<td>2.0282</td>
<td>5.6768</td>
</tr>
</tbody>
</table>

**Table 5. Simulation ($\tau = 30$) descriptive results**

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Misclassification cost for class</th>
<th>Population size</th>
<th>Mean in %</th>
<th>$aSTDEV$ in %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>neg. pos.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>0.1 0.9</td>
<td>13</td>
<td>7.6683</td>
<td>6.5102</td>
</tr>
<tr>
<td>SVM</td>
<td>0.3 0.9</td>
<td>27</td>
<td>8.3522</td>
<td>12.2889</td>
</tr>
<tr>
<td>SVM</td>
<td>0.5 0.9</td>
<td>96</td>
<td>5.2979</td>
<td>9.4553</td>
</tr>
<tr>
<td>SVM</td>
<td>0.9 0.9</td>
<td>155</td>
<td>4.3800</td>
<td>7.9053</td>
</tr>
<tr>
<td>SVM</td>
<td>0.9 0.1</td>
<td>361</td>
<td>3.1766</td>
<td>7.4125</td>
</tr>
<tr>
<td>LONG</td>
<td>- -</td>
<td>423</td>
<td>2.8080</td>
<td>6.9184</td>
</tr>
</tbody>
</table>

The population means basically confirm above classification results. Those cases, where the class “positive” is assigned very high misclassification costs, reveal highest returns. The results therefore provide first evidence that the SVM-based trading strategy outperforms the benchmark “all-long strategy”.

To further statistically explore this, corresponding null- and alternative hypotheses are formulated:

$$H_0 : \mu(R_{SVM}) \leq \mu(R_{LONG}) \quad \text{and} \quad H_1 : \mu(R_{SVM}) > \mu(R_{LONG})$$

The respective SVM return population is termed $R_{SVM(cost \ “negative”; \ cost \ “positive”)}$ depending on the applied misclassification cost input.
If a null hypothesis can be rejected, we can statistically corroborate a higher population mean of \( R_{\text{SM (neg. pos.)}} \) compared to \( R_{\text{LONG}} \) for that classification cost setting at a given level of significance. We conduct two-sample \( t \)-tests assuming unequal variances with a hypothesized mean difference of zero. Test statistics are summarized in Table 6.

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>( \tau )</th>
<th>( t )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu(R_{\text{SM (1.0/9.0)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>15</td>
<td>2.71 **</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.3/0.9)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>15</td>
<td>2.59 ***</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.5/0.9)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>15</td>
<td>2.72 ***</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.9/0.9)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>15</td>
<td>2.16 **</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.9/0.3)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>15</td>
<td>0.80</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (1.0/9.0)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>30</td>
<td>2.55 **</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.3/0.9)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>30</td>
<td>2.28 **</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.5/0.9)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>30</td>
<td>2.42 ***</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.9/0.9)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>30</td>
<td>2.18 **</td>
</tr>
<tr>
<td>( \mu(R_{\text{SM (0.9/0.3)}} ) \leq \mu(R_{\text{LONG}}) )</td>
<td>30</td>
<td>0.71</td>
</tr>
</tbody>
</table>

*** / ** indicate significance at the 1% / 5%-level.

Null hypotheses can be rejected at high levels of significance for those cases where high / medium class “positive” precision is achieved by applying above described post-processing steps, i.e. for high / medium class “positive” misclassification costs. In other words, the results provide evidence that the SVM-based forecasting model works better than the highly competitive benchmark. Overall, it may therefore be concluded that the proposed SVM text mining approach is capable of identifying (25%) riskiest events.

6. Summary and conclusion

Traditional risk management tools such as stress testing are capable of quantifying the risk exposure associated with potential critical market events. Management implications of this generalized hypothetical risk exposure are, however, rather unclear. While critical market events appear infrequently, every counteraction to take account of these events would need to be implemented on a continuous basis. It is therefore already valuable to be able to assess the risk exposure that results from unanticipated critical events on an intraday real-time basis. In order to discover and consequently react to intraday risk exposures promptly, all available information should be taken account of. Traditional risk management tools, however, usually neglect one of the largest sources of information, i.e. unstructured qualitative data.

It is against this background that we propose a state-of-the-art text mining approach that incorporates the valuable source of unstructured qualitative data on a real-time intraday basis. By means of an empirical study, we show that today’s technology is capable of extracting valuable information from corporate disclosures for risk management purposes. Both a “classic” and a newly developed domain-specific “simulation-based” evaluation metric confirm the suitability of our approach to identify most critical, i.e. volatility-enhancing, market events. We therefore conclude intraday market risk exposures can be discovered utilizing text mining techniques.

Future research will include the application of further text mining techniques (e.g. Neural Networks, k-Nearest Neighbour) to an extended dataset. It is, for example, of greatest interest whether or not the proposed approach also works in times of market turmoil.

7. References


