An intraday market risk management approach based on textual analysis

Sven S. Groth [a], Jan Muntermann [b],*

[a] E-Finance Lab, House of Finance, Grüneburgplatz 1, 60323 Frankfurt, Germany
[b] Goethe-University Frankfurt, House of Finance, Grüneburgplatz 1, 60323 Frankfurt, Germany

ARTICLE INFO

ABSTRACT

The management of financial risk is one of the most challenging tasks of financial institutions. In the last two decades, diverse quantitative models and approaches have been developed and refined to address the impact of volatile markets on business. Whereas existing approaches have intensively utilized structured data such as historical price series, little attention has been paid to unstructured (textual) data, which could be a large source of information in this context. Previous empirical research has shown that certain news stories, such as corporate disclosures, can cause abnormal price behavior subsequent to their publication. On the basis of a data set comprising such news stories as well as intraday stock prices, this paper explores the risk implications of information being newly available to market participants. After showing that such events can significantly drive stock price volatilities, this research aims at identifying among the textual data provided those disclosures that have resulted in most supranormal risk exposures. To this end, four different learners — Naïve Bayes, k-Nearest Neighbour, Neural Network, and Support Vector Machine — have been applied in order to detect patterns in the textual data that could explain increased risk exposure. Two evaluations are presented in order to assess the learning capabilities of the approach in the context of risk management. First, "classic" data mining evaluation metrics are applied and, second, a newly developed simulation-based evaluation method is presented. Evaluation results provide strong evidence that unstructured (textual) data represents a valuable source of information also for financial risk management — a domain in which, in the past, little attention has been paid to unstructured data. With regard to classification performance, it is also shown that there exist significant differences between the applied learning techniques.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Financial modeling of market risks, i.e. the management of losses due to movements in financial market prices, has been a subject of research for the last few decades. Today, traditional financial risk management tools, such as Value-at-Risk and Stress Testing, make use of quantitative data stored in structured databases [37]. While the approaches to analyze such structured data, e.g. historical price series, have been continuously improved (for an example of intraday effects see [25]) , little attention has been paid in this context to the analysis of unstructured qualitative data.

Especially when assessing intraday market risk exposures resulting from market events, such as critical corporate disclosures that were not anticipated by market participants, there exists limited quantitative data (at event time) that could be analyzed in such situations.

However, the disclosures contain qualitative data representing a potential source of information that is not taken into consideration by traditional risk management tools. Nonetheless, the management of intraday market risk is a challenge for different market participants

engaged in frequent trading, such as high frequency traders, floor traders, and market makers [25]. Also, using quantitative intraday data to forecast intraday volatilities is still at its infancy and quite unsuitable to address event risks as described before [39]. Event risk plays an important role, especially in the case of small time horizons when relevant information about a company is newly available to the market. In this context, Campbell et al. [10] have observed that firm-level variance has more than doubled in the last three decades while market and industry variances have remained stable over that period.

Therefore, the goal of this paper is to explore how existing risk management approaches can be supported by utilizing unstructured textual data. In contrast to the wide range of publications that focus on quantitative risk management approaches, we will focus on the potentials of qualitative data and the corresponding data analysis methodologies.

The remainder of our paper is structured as follows: In Section 2, we provide a literature review on related work in the fields of financial risk management and on text mining in the context of financial forecasting. Furthermore, a theoretical foundation is provided. Then, Section 3 illustrates how and why we have selected our data set. In Section 4, we present a text mining approach to identify significant risk exposures that result from textual information newly available to market participants. In Sections 5 and 6 we provide evaluations of our

* Corresponding author. Fax: +49 69 798 33910.
  E-mail address: muntermann@wiwi.uni-frankfurt.de (J. Muntermann).

approach. We present both a "classic" evaluation based on traditional evaluation metrics (Section 5) and a domain-specific evaluation on the basis of a simulation (Section 6). In Section 7, we conclude with a summary and an outlook on further research.

2. Related work

2.1. Management of market risks

Today, financial market risks are modeled with the help of 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. 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 [33]. Here, risk positions are assessed for longer holding periods such as 10 days. In order to address intraday market risks, recent research aims to extend traditional Value-at-Risk to cover shorter (intraday) horizons [19]. However, due to the used confidence interval, Value-at-Risk does not provide support for managing extraordinary losses, and since the underlying calculations are based on historical or artificial data sets, intraday losses that could result from critical market events are not covered [2]. Another established risk management approach that has a focus on events that could result in extraordinary losses, is stress testing. Utilizing historical or artificial quantitative data, stress testing aims at supporting managers to assess financial consequences of critical market behavior, i.e. to conduct what-if-analyses for such scenarios [32]. Central objective of these approaches is to quantify hypothetical risk exposures for different scenarios, but not to discover actual intraday risk exposures that result from given but unforeseen critical events. Utilizing only quantitative data seems inappropriate when there is also unstructured data available, which is indeed more complex to process and to analyze but which, however, could contain valuable information.

2.2. "Financial" text mining

An appearance of price volatility undoubtedly constitutes a potential risk. The fact that there is already a multitude of research on how to measure volatility underpins this perception [44]. Nonetheless, there exists only little research, if any at all — depending on the definition — that aims at utilizing unstructured data sources in the context of risk management by applying text mining techniques. However, volatility forecasting is highly relevant for financial risk management, especially with a focus on short periods [15]. While there are quantitative approaches for intra-day volatility forecasts [19,39], there exists little research on utilizing unstructured data in this field.

Even though the data/text mining approaches by [47,51] are classified as volatility forecasting systems by [42], we believe that this is the case only within strict limitations: [47] aim at forecasting abnormal stock returns, i.e. not volatility. The rule-based hand-crafted classifier by [51] does not reflect our understanding of text mining (see below, [28]).

Another line of thought 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. [57] present one of the first applications of text mining techniques addressing financial forecasting problems. Later work focuses 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 [26,30,41,47,48]. Several IT artifacts that were developed in order to provide risk management functionalities [38] or to exploit text mining [23,24,41] in the domain of stock trading have been presented in the past.

On the basis of a multi-agent approach, Luo et al. [38] present an IT infrastructure that provides decision support for stock trading. The features of the risk management agent are mainly based on quantitative data (such as prices and trading volumes), an approach that has been the state-of-the-art in financial research for several years [3,8]. With regard to unstructured data, "breaking news" related to given shares are monitored. While the presented infrastructure provides functionalities to monitor news releases and to map those to given securities, there are no functionalities that aim at assessing the corresponding risk exposures. In [23,24,41], system architectures are presented that aim at estimating stock price trends (positive, negative, or neutral) on the basis of textual data. Having a focus on implementing profitable trading strategies, no information on volatility increases is provided, which would be relevant to support risk management. Consequently, since they focus on trend forecasts and trading strategies, existing system concepts provide only little evidence that unstructured (textual) data can be utilized to support the management of market risks.

2.3. Theoretical background

At an efficient capital market, prices will promptly adjust to information newly available to market participants. This well-established efficient market hypothesis EMH (semi-strong-form) has been developed by Eugene Fama and goes back to the 1960s [21,22]. In financial research, EMH provides the theoretical basis for various research approaches, among them event study analysis. Here, price adjustments to previously unknown events are statistically analyzed, such as the speed at which prices fully reflect new information available. Furthermore, the information content of different news types can be explored by statistically proving the existence of abnormal price behavior following a specific event type. Significant information content has been proven to exist on the basis of abnormal stock returns for corporate disclosures such as announcements on E-Commerce initiatives and IT investments [18,50]. Abnormal market behavior has also been observed, from a more general perspective, for corporate disclosures that were published due to regulatory legislation [12,43]. Most event study analyses are based on abnormal returns, i.e. returns observed subsequent to an event and adjusted by a market trend that has, for example, been calculated with a market model [11].

In the context of market risk management, significant price volatility represents a relevant measure to focus on [25]. Several papers in finance literature have detected no or only a weak connection between the volatility of stock prices and news releases such as public disclosures of monetary policy decisions [5]. These results provide evidence that a focus on specific news types is appropriate when aiming at analyzing news content carrying information that could result in significant risk exposure.

3. Data set and risk modeling

As a basis for our research, we use a data set that is composed of news and stock price data. The selected news data appears suitable in the context of market risk management, since we aim at finding such news stories that are associated with highest market risk. We assume that the type of news we have selected is appropriate for such an endeavor, which we will illustrate in the following. High-frequency intraday stock price data are associated with the news data to analyze potential relations.

3.1. Data set: corporate disclosures and stock prices

From a market risk management perspective, one should be interested in primary information sources, i.e. on sources of information that are expected to significantly drive market prices. We therefore